

ФГАОУ ВПО «Уральский федеральный университет имени первого
Президента России Б.Н. Ельцина»

На правах рукописи



Копосов Александр Сергеевич

**МАТЕМАТИЧЕСКОЕ И АЛГОРИТМИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДЛЯ
ОБРАБОТКИ СЛУЧАЙНЫХ ДАННЫХ С ОГРАНИЧЕННОЙ
ОБЛАСТЬЮ РАССЕЯНИЯ**

Специальность 05.13.01 – Системный анализ, управление и обработка
информации (связи и информатизация) (технические науки)

Диссертация на соискание ученой степени
кандидата технических наук

Научный руководитель:
д.т.н., проф. Поршнева
Сергей Владимирович

Екатеринбург – 2016

СПИСОК СОКРАЩЕНИЙ

ГА – генетический алгоритм

ДРОСП – двумодальное распределение ограниченной случайной последовательности

НР - нормальное распределение

ПР – плотность распределения

СНР стандартное нормальное распределение

СП – случайная последовательность

СПООР – случайная последовательность с ограниченной областью рассеяния

УНР – усеченное нормальное распределение

ФР – функция распределения

ОГЛАВЛЕНИЕ

| | |
|--|----|
| ВВЕДЕНИЕ | 5 |
| 1 АНАЛИЗ СОСТОЯНИЯ ПРЕДМЕТНОЙ ОБЛАСТИ. ПОСТАНОВКА ЗАДАЧ ИССЛЕДОВАНИЯ | 12 |
| 1.1 Анализ особенностей распределений случайных последовательностей с ограниченной областью рассеяния | 12 |
| 1.2 Постановка задачи оценки функций распределения случайных последовательностей с ограниченной областью рассеяния и анализ методов ее решения | 39 |
| 1.3 Постановка задач исследования | 47 |
| 2 РАЗРАБОТКА МАТЕМАТИЧЕСКОГО И АЛГОРИТМИЧЕСКОГО ОБЕСПЕЧЕНИЯ ДЛЯ ОЦЕНИВАНИЯ ПАРАМЕТРОВ ОДНОМОДАЛЬНЫХ РАСПРЕДЕЛЕНИЙ СЛУЧАЙНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ С ООР | 49 |
| 2.1 Постановка задачи оценивания параметров одномодальных распределений случайных последовательностей с помощью ГА | 49 |
| 2.2 Исследование особенностей одномодальных распределений случайных последовательностей на основе статистического моделирования | 54 |
| 2.3 Анализ точности оценок параметров одномодальных распределений случайных последовательностей с ограниченной областью рассеяния с помощью ГА | 64 |
| 2.4 Методика оценивания параметров одномодальных распределений случайных последовательностей с ограниченной областью рассеяния, основанный на использовании ГА | 69 |
| 2.5 Выводы по главе | 70 |
| 3 РАЗРАБОТКА МАТЕМАТИЧЕСКОГО И АЛГОРИТМИЧЕСКОГО ОБЕСПЕЧЕНИЯ ДЛЯ ОЦЕНИВАНИЯ ПАРАМЕТРОВ ДВУМОДАЛЬНЫХ РАСПРЕДЕЛЕНИЙ СЛУЧАЙНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ С ООР | 71 |
| 3.1 Постановка задачи оценивания параметров двумодальных распределений случайных последовательностей с помощью ГА | 71 |
| 3.2 Анализ точности оценивания параметров двумодальных распределений, получаемых в соответствии с алгоритмом, обоснованным в Главе 2 | 72 |
| 3.3 Исследование особенностей численных оценок параметра размытости в аппроксимации Розенблатта-Парзена | 78 |
| 3.4 Использование аппроксимации Розенблатта-Парзена для восстановления функции распределения дискретной случайной величины | 83 |
| 3.5 Использование аппроксимации Розенблатта-Парзена для восстановления функции распределения непрерывной случайной величины с ограниченным одномодальным законом распределения | 87 |

| | | |
|-----|---|-----|
| 3.6 | Методика оценивания параметров случайной величины со смешанным двумодальным законом распределения на основе совместного использования аппроксимации Розенблатта-Парзена, метода мнимых источников и ГА..... | 91 |
| 3.7 | Итерационный алгоритм оценивания параметров случайной величины с двумодальным законом распределения на основе использования аппроксимации Розенблатта-Парзена, метода мнимых источников и ГА | 96 |
| 3.8 | Выводы по главе | 100 |
| 4 | ОПИСАНИЕ ПРОГРАММНОЙ БИБЛИОТЕКИ «ES&RP», ОБЕСПЕЧИВАЮЩЕЙ ПРАКТИЧЕСКОЕ ИСПОЛЬЗОВАНИЕ МАТЕМАТИЧЕСКОГО И АЛГОРИТМИЧЕСКОГО ОБЕСПЕЧЕНИЯ ДЛЯ ОЦЕНКИ ФР И ПВ СПООР..... | 101 |
| 4.1 | Назначение программной библиотеки | 101 |
| 4.2 | Используемые технические средства | 101 |
| 4.3 | Описание модулей программной библиотеки | 101 |
| 4.4 | Примеры использования программной библиотеки | 116 |
| 4.5 | Выводы по главе | 120 |
| | ЗАКЛЮЧЕНИЕ..... | 121 |
| | СПИСОК ЛИТЕРАТУРЫ..... | 122 |
| | ПРИЛОЖЕНИЕ А. ПРИМЕРЫ ФР И ПР ДЛЯ РАЗЛИЧНЫХ ПАРАМЕТРОВ УНР..... | 130 |
| | ПРИЛОЖЕНИЕ Б. РЕЗУЛЬТАТЫ ПРИМЕНЕНИЯ ИТЕРАЦИОННОГО АЛГОРИТМА К ВОССТАНОВЛЕНИЮ ПР СЛУЧАЙНЫХ ВЕЛИЧИН С ДВУМОДАЛЬНЫМ РАСПРЕДЕЛЕНИЕМ С ОГРАНИЧЕННОЙ ОБЛАСТЬЮ РАССЕЯНИЯ | 133 |
| | ПРИЛОЖЕНИЕ В. ПРИМЕНЕНИЕ МЕТОДА МНИМЫХ ИСТОЧНИКОВ И АППРОКСИМАЦИИ РОЗЕНБЛАТТА-ПАРЗЕНА ДЛЯ АНАЛИЗА СЕТЕВОГО ТРАФИКА | 138 |
| | ПРИЛОЖЕНИЕ Г. РЕЗУЛЬТАТЫ ОЦЕНКИ ПАРАМЕТРОВ РАСПРЕДЕЛЕНИЙ ПО КАЖДОМУ ИЗ ДНЕЙ, ПОЛУЧЕННЫЕ ПО ПОЛНОЙ ВЫБОРКЕ АГРЕГИРОВАННЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ В РАЗРЕЗЕ КОЛИЧЕСТВА ПАКЕТОВ..... | 165 |
| | ПРИЛОЖЕНИЕ Д. РЕЗУЛЬТАТЫ ОЦЕНКИ ПАРАМЕТРОВ РАСПРЕДЕЛЕНИЙ ПО КАЖДОМУ ИЗ ДНЕЙ, ПОЛУЧЕННЫЕ ПО ПОЛНОЙ ВЫБОРКЕ АГРЕГИРОВАННЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ В РАЗРЕЗЕ РАЗМЕРА ПОТОКОВ..... | 171 |
| | ПРИЛОЖЕНИЕ Е. РЕЗУЛЬТАТЫ ВЫЧИСЛЕНИЯ ПАРАМЕТРОВ РАСПРЕДЕЛЕНИЙ СЛУЧАЙНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ, СОДЕРЖАЩИХ ЗНАЧЕНИЯ РАЗМЕРОВ ПАКЕТОВ | 177 |
| | ПРИЛОЖЕНИЕ Ж. АКТЫ О ВНЕДРЕНИИ | 180 |
| | ПРИЛОЖЕНИЕ З. СВИДЕТЕЛЬСТВО О РЕГИСТРАЦИИ ПРОГРАММЫ ДЛЯ ЭВМ... | 184 |

Введение

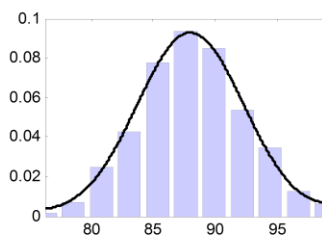
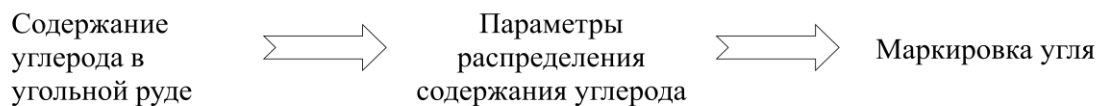
Актуальность темы диссертационного исследования определяется непрерывно возрастающими требованиями к достоверности обработки и интерпретации технологической информации для управления производственными процессами и качеством (проблема Statistical Quality Control).

С математической точки зрения задача обработки технологической информации относится к задаче оценивания функций распределений (ФР) и плотностей вероятностей (ПВ) эмпирических зависимостей по некоторой случайной выборке данных. Для решения этой задачи сегодня разработано большое число различных методов, основанных на подходах параметрической и непараметрической статистики. Однако, только относительно небольшая часть данных методов доведена до законченных программных реализаций, включенных в виде соответствующих инструментов в современные математические пакеты (MATLAB, Statistica, Statgraphics, Mathematica, Maple и др.) и программные библиотеки ряда языков программирования (C++, Python, R и др.)

Анализ систем, используемых в различных отраслях промышленности (угледобывающей, деревообрабатывающей, нефте- и газотранспортной, связи, приборостроительной и др.), показывает, что порождаемая ими технологическая информация с математической точки зрения представляют собой случайные числа, область рассеяния которых является ограниченной (например, время работоспособности приборов, диаметры бревен, концентрации углерода в углях, число пакетов, передаваемых в компьютерных сетях, в течении выбранного временного интервала и др.).

Приведем некоторые примеры подобных задач.

1. Определение параметров распределения содержания углерода в коксующихся углях при торгово-промышленной маркировке угля, которое, исходя из объективной оценки, ограничено снизу примерно 70% и сверху 100%. Данный факт обусловлен тем, что на месторождениях тех или иных количества добывают угли с разной степенью углефикации в указанных пределах. При добыче, отгрузке и транспортировке угля происходит его перемешивание, что может рассматриваться как случайный процесс в ограниченной области рассеяния. Затем происходит многократный забор проб угля. В итоге, получают случайную последовательность, состоящую из значений процентов содержания углерода в пробах угольной руды. На основании параметров данного распределения происходит маркировка угля, т.е. определяется цена угольной руды. Принимая во внимание, что содержание углерода в угле опеределает его сортность и, следовательно, цену, понятно, что решение данной задачи имеет важное экономическое значение (рисунок В.1).



| Марки угля | Буквенное обозначение марок | содержание углерода С, % |
|----------------------|-----------------------------|--------------------------|
| Бурые | Б | 76 и менее |
| Длиннопламенные | Д | 76 |
| Газовые | Г | 83 |
| Жирные | Ж | 86 |
| Коксовые | К | 88 |
| Отощённо-спекающиеся | ОС | 89 |
| Тошние | Т | 90 |
| Антрациты | А | 91 и более |

Рисунок В.1. К объяснению технологии торгово-промышленной маркировки углей

2. Торгово-промышленная сортировка древесного сырья в деревообрабатывающей промышленности, в частности, сортировка бревен по диаметрам, точность которой определяет эффективность использования древесины при оцилиндровке, и, следовательно, лесопильного производства в целом. Данная сортировка проводится на основе информации о значениях диаметров комлей деревьев данного штабеля (рисунок В.2).

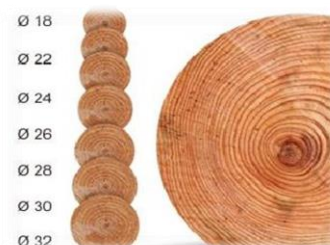
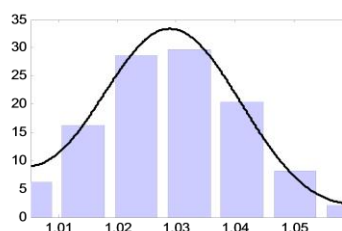
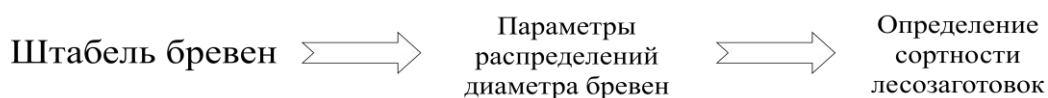


Рисунок В.2. К объяснению технологии торгово-промышленной маркировки лесоматериалов

3. Оценивание прочностной надежности элементов и объектов нефтегазового оборудования с целью определения вероятности безотказной работы нефтепроводов, зависящей от напряжения в трубопроводе и предельного напряжения в трубопроводе, которые также являются случайными величинами с ограниченной областью рассеяния (ограничены слева нулем). При наличии оценки вероятности безотказной работы оказывается возможным вычислить ресурс работы нефтепровода и оценить безопасность технических систем (рисунок В.3).

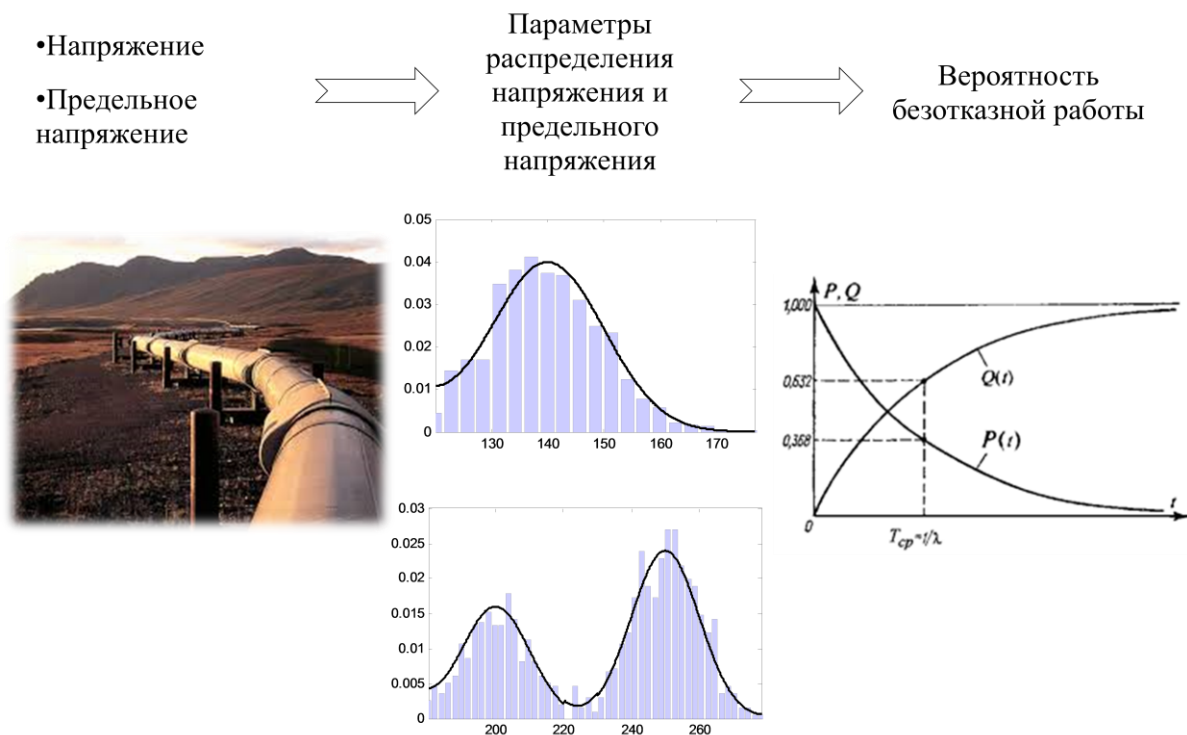


Рисунок В.3. К постановке задачи оценивания вероятности безаварийной работы нефтепровода

4. Оценивание статистических свойств информационных потоков, передаваемых в телекоммуникационных сетях, важность которой обусловлено постоянно увеличением объемов трафика и типов устройств, генерирующих трафик (рисунок В.4).

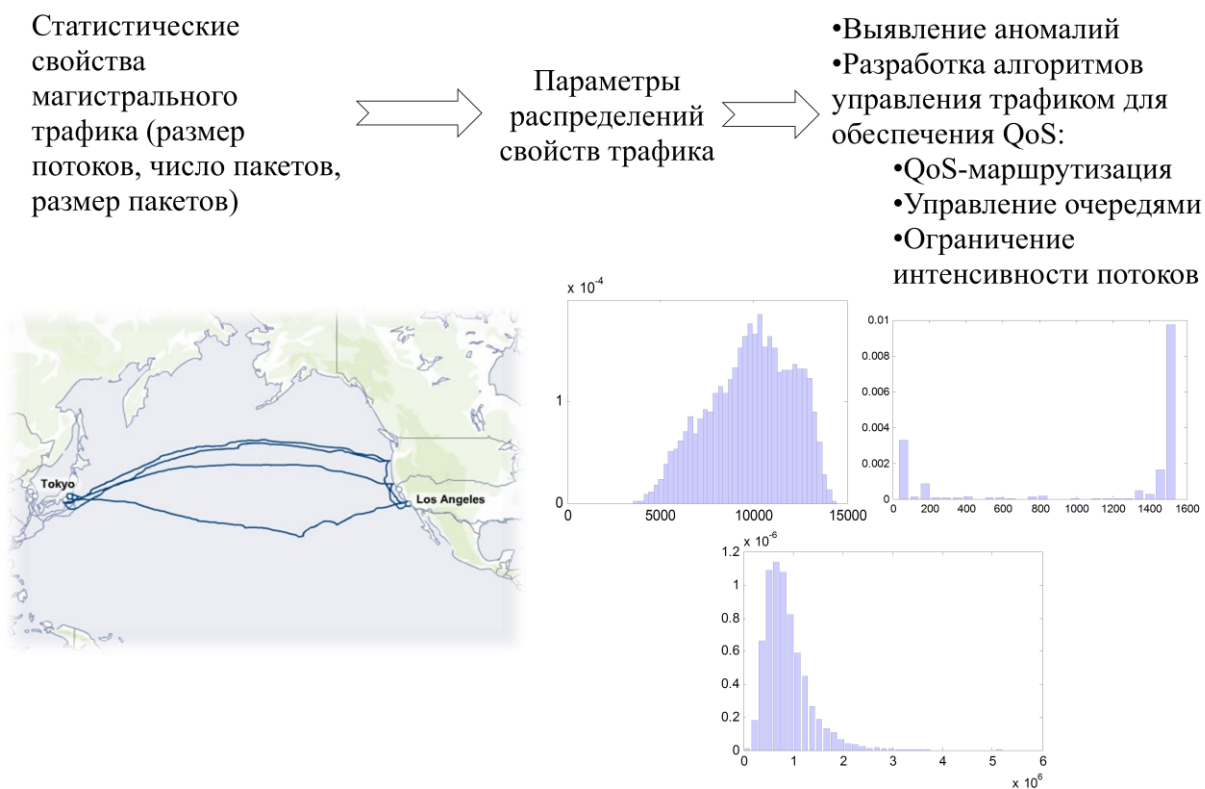


Рисунок В.4. К постановке задачи анализа статистических свойств Интернет-трафика

Здесь для эффективного проектирования телекоммуникационных сетей нового поколения необходимо понимать особенности информационных потоков и в соответствии с ними использовать адекватные алгоритмы эффективного управления потоками трафика, например: QoS-маршрутизация, управление очередями и планирования обслуживания пакетов, ограничение интенсивности потоков и др.

Теория случайных процессов с ограниченной областью рассеяния (ООР), порождающих соответствующие выборки случайных величин, была построена в работах А. Эйнштейна и М. Смолуховского, в которых они предложили математические модели ФР и ПР случайных величин данного типа. Однако при оценке параметров этих моделей по экспериментальным данным возникает необходимость решения той или иной системы нелинейных уравнений с помощью какого-либо известного итерационного метода (метод Ньютона, симплекс-метод и т.д. [2,5,6,13,14,27,44,71,76,78,102,112,116]). Их сходимость, как известно [30,84,108,109,110], определяется удачным выбором начального приближения, сделать который в многомерном случае оказывается не всегда возможным. Данное обстоятельство, с нашей точки зрения, явилось одной из причин недостаточного широкого использования моделей Эйнштейна и Смолуховского в задачах обработки информации.

В то же время сегодня разработаны методы решения экстремальных задач, основанные на случайном поиске (генетические алгоритмы, муравьиные алгоритмы и т.д.), сходимость которых к истинному решению, как утверждается, не зависит от начального приближения [16,22,85]. Однако, не существует алгоритмов их использования в задаче аппроксимации эмпирических зависимостей и распределений случайных величин с ООР.

В этой связи оказывается актуальной разработка научно-обоснованных алгоритмов оценивания параметров ФР и ПВ случайных величин с ООР с помощью генетических алгоритмов (ГА), объединяющих параметрические и непараметрические методы оценивания, и их программных реализаций.

Объектом исследования являются методы анализа и обработки технологической информации.

Предмет исследования

Алгоритмическое и математическое обеспечение для оценивания ФР и ПР случайных величин с ООР.

Целью работы является разработка математического и алгоритмического обеспечения для обработки технологической информации, представляющую собой СПООР, на основе использования методов параметрической и непараметрической статистики и эвристических методов.

Для достижения поставленной цели сформулированы и решены следующие **основные задачи исследования**:

1. Анализ особенностей одномодальных распределений СПООР на основе статистического моделирования.

2. Разработка математического и алгоритмического обеспечения для оценивания параметров распределений СПООР, обеспечивающего:

- оценку параметров одномодальных распределений с ООР на основе ГА;
- оценку параметров двумодальных случайных величин с ООР на основе ГА и аппроксимации Розенблатта-Парзена.

3. Разработка программной библиотеки, обеспечивающей практическое использование математического и алгоритмического обеспечения для оценивания параметров ФР и ПВ случайных величин с ООР.

Методы исследования

В работе использованы методы математической статистики, вычислительной математики, системного анализа, имитационного моделирования, ГА.

Теоретико-методологической базой исследований, проведенных в диссертационной работе, послужили труды В.В. Курейчика, Л.А. Гладкова, Х. Гулд, Я. Тобочника, А.Н. Колмогорова, В.Р. Матвиевский, А.И. Орлова, В.А. Симахина, С.Л. Голофаства, А. Эйнштейна, М. Смолуховского, Э. Парзена, М. Розенблатта.

Результаты, выносимые на защиту:

1. Научно-обоснованное математическое и алгоритмическое обеспечение для оценивания ФР и ПВ случайных величин с ООР, включающая в себя:

- методику оценки параметров одномодальных распределений с ООР на основе ГА;
- комплексную методику оценки параметров двумодальных случайных величин с ООР на основе ГА и аппроксимации Розенблатта-Парзена.

Соответствует п. 4 паспорта специальности: Разработка методов и алгоритмов решения задач ... обработки информации; п 5: Разработка специального математического и алгоритмического обеспечения систем ... обработки информации.

2. Алгоритм нахождения оптимального значения параметра размытости h^* аппроксимации Розенблатта-Парзена, адаптированный для дискретных целочисленных случайных последовательностей. *Соответствует п. 4 паспорта специальности: Разработка методов и алгоритмов решения задач ... обработки информации; п 5: Разработка специального математического и алгоритмического обеспечения систем ... обработки информации.*

3. Алгоритм нахождения оптимального значения параметра размытости h^* аппроксимации Розенблатта-Парзена, учитывающий особенности машинного нуля при вычислении на ЭВМ. *Соответствует п. 4 паспорта специальности: Разработка методов и алгоритмов решения задач ... обработки информации; п 5 паспорта специальности: Разработка*

специального математического и алгоритмического обеспечения систем ... обработки информации.

4. Программный пакет, реализующий математическое и алгоритмическое обеспечение для оценивания ФР и ПВ случайных величин с ООР. *Соответствует п 5 паспорта специальности: Разработка специального математического и алгоритмического обеспечения систем ... обработки информации*

Научную новизну исследования представляют следующие результаты исследования:

1. Научно-обоснованное математическое и алгоритмическое обеспечение для оценивания ФР и ПВ случайных величин с ООР, включающая в себя:

- методику оценки параметров одномодальных распределений с ООР на основе ГА;
- комплексную методику оценки параметров двумодальных случайных величин с ООР на основе ГА и аппроксимации Розенблатта-Парзена.

2. Алгоритм нахождения оптимального значения параметра размытости h^* аппроксимации Розенблатта-Парзена, адаптированный для дискретных целочисленных случайных последовательностей.

3. Алгоритм нахождения оптимального значения параметра размытости h^* аппроксимации Розенблатта-Парзена, учитывающий особенности машинного нуля при вычислении на ЭВМ.

Теоретическая значимость представляют следующие результаты работы:

1. Модифицированная для целочисленных последовательностей формула для информационного функционала $\varphi(h)$, используемого в аппроксимации Розенблатта-Парзена, в которой решена проблема немонотонности $\varphi(h)$, что обеспечивает более точную оценку оптимального значения параметра размытости h .

2. Результаты исследований особенностей случайных блужданий в ООР.

3. Формулировки задачи оценивания параметров одномодальных и двумодальных распределений с ООР в терминах ГА.

Практическая значимость исследования

1. Обоснованы рекомендации по выбору областей поиска значений параметров одномодальных распределений с ООР с помощью ГА.

2. Обоснованы рекомендации по выбору областей поиска значений параметров двумодальных распределений с ООР с помощью ГА и аппроксимации Розенблатта-Парзена.

3. Разработана программная библиотека, обеспечивающая практическое использование математического и алгоритмического обеспечения для оценки ФР и ПВ СПООР, на которую получено

свидетельство о государственной регистрации программы для ЭВМ № 2016614275.

Полученные в диссертации результаты используются в ООО «Эйконикс», Уральский Банк ПАО «Сбербанк России», ОАО «АйТи-Мед», АО «ЭР-Телеком Холдинг». Основные результаты работы внедрены в учебный процесс Института радиоэлектроники и информационных технологий Уральского федерального университета им. первого Президента России Б.Н. Ельцина при подготовке бакалавров и магистров по направлению «Информационные системы и технологии»

Апробация работы. Материалы работы докладывались на следующих научных конференциях: Международной научно-практической конференции: «Инновационное развитие: физико-математические и технические науки», Москва, 28 августа 2014 г.; Международной научно-практической конференции «Новое слово в науке: перспективы развития», Чебоксары, 10 сентября 2014 г.; Международной IEEE Сибирская конференция по управлению и связи SIBCON-2015, Омск, 21-23 мая 2015 г.; Международной Крымской конференции «СВЧ-техника и телекоммуникационные технологии» (КрыМиКо'2015), Севастополь, 6-12 сентября 2015; Международной IEEE конференции AICT, Ростов-на-Дону, 14-16 октября 2015 г.; Международная научная конференция «Современные методы прикладной математики, теории управления и компьютерных технологий», Воронеж, 20-26 сентября 2016 г.

Публикации. По теме диссертации опубликовано 19 работ, отражающих основные положения исследования, среди которых 11 статей в журналах, рекомендованных ВАК РФ, из них одна статья вышла в переводной версии журнала, 7 текстов докладов в материалах международных научно-практических конференций. Получено свидетельство о государственной регистрации программы. В работах, опубликованных в соавторстве, лично соискателю принадлежат: исследование особенностей случайных блужданий в ООР; разработка математического и алгоритмического обеспечения для оценивания параметров ФР и ПР с помощью генетических алгоритмов и аппроксимации Розенблатта-Парзена.

Структура и объем работы

Диссертация состоит из введения, четырех глав, заключения, списка литературы из 127 наименований, 8 приложений, содержит 67 рисунков и 39 таблиц. Основной текст работы составляет 129 страниц, общий объем – 184 страницы.

Equation Chapter (Next) Section 1

1 Анализ состояния предметной области. Постановка задач исследования

1.1 Анализ особенностей распределений случайных последовательностей с ограниченной областью рассеяния

1.1.1 Анализ особенностей одномодальных распределения случайных последовательностей с ограниченной областью рассеяния

Нормальное распределение (НР), зависящее от двух параметров μ, σ ($N(\mu, \sigma)$), является одним из наиболее широко применяемых на практике распределений, что связано с его известной универсальностью. Свойства НР были исследованы в работах Муавра, Лапласа, Гаусса, Чебышева, Ляпунова, Бернштейна [7,15,18,19,37,38,40,73,80,81]. Теоретическим обоснованием применимости НР является центральная теорема Ляпунова, имеющая следующую формулировку [12,20,51]:

Теорема Ляпунова: Пусть $X_1, X_2 \dots X_n$ – последовательность взаимно независимых случайных величин. Обозначим

$$\mu_k = E[X_k], \sigma_k^2 = D[X_k], B_n^2 = D\left[\sum_{k=1}^n X_k\right].$$

Если существует положительное число $\delta > 0$, такое что

$$\lim_{n \rightarrow \infty} \left(\frac{1}{B_n^{2+\delta}} \cdot \sum_{k=1}^n E|X_k - a_k|^{2+\delta} \right) = 0,$$

то

$$\lim_{n \rightarrow \infty} \left(P \left\{ \frac{1}{B_n^{2+\delta}} \cdot \sum_{k=1}^n E|X_k - a_k|^{2+\delta} < x \right\} \right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-z^2/2} dz.$$

Из теоремы Ляпунова следует, что распределение суммы независимых случайных величин с любым исходным законом распределения будет нормальным, если число слагаемых достаточно велико, а вклад каждого в сумму мал.

Перечислим, следуя [1] причины, которые обеспечили НР центральное место в математической статистике:

- природа возникновения случайных чисел с нормальным законом распределения аналогична природе многих физических процессов, порождающих результаты обрабатываемых наблюдений;
- в соответствии с центральной предельной теоремой Ляпунова при возрастании объема выборки в пределе большинство распределений можно с достаточной точностью аппроксимировать НР;
- НР легко нормируется, аппроксимируется, а также обладает свойством аддитивности.

1.1.1.1 Нормальное распределение с неограниченной областью рассеяния

Нормальное распределение связано с уравнением Фоккера-Планка

$$\frac{\partial}{\partial t} f(t, x) = D \frac{\partial^2}{\partial x^2} f(t, x), \quad (1.1)$$

описывающего процесс движения точечной частицы под действием случайных сил вдоль бесконечной прямой под действием случайных толчков (броуновская частица),

где

$$f(t, x) - \text{плотность распределения частиц } f(t, x) = \frac{\partial}{\partial x} P(t, x),$$

$P(t, x)$ – вероятность нахождения частицы в момент времени t в точке x ;

D – коэффициент диффузии.

Решение уравнения (1.1) для неограниченной области рассеивания броуновских частиц было найдено А. Эйнштейном [82]

$$f(t, x) = \frac{A}{\sqrt{4\pi D}} \frac{e^{-\frac{x^2}{4Dt}}}{\sqrt{t}}, \quad (1.2)$$

здесь A – нормировочный коэффициент, выбираемый из условия $\int_{-\infty}^{\infty} f(t, x) dx = 1$.

Выражение (1.2) связано с НР. Действительно, трактуя величину $2Dt$ как дисперсию случайного процесса для конечного времени t , приходим к гауссовой ПР $f(x, 0, \sigma) = N(0, \sigma)$:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{x^2}{2\sigma^2}}. \quad (1.3)$$

Отметим, что аналогичный результат можно получить, рассмотрев броуновское движение точечной частицы на бесконечной прямой под действием случайных толчков [20], вызывающих с вероятностью p ее смещение в одну сторону, и с вероятностью q ее смещение в другую сторону. Тогда вероятность $P_n(m)$ того, что после n случайных толчков частица окажется в m -ой точке рассчитывается по формуле

$$P_n(m) = \frac{n!}{\left(\frac{n+m}{2}\right)! \left(\frac{n-m}{2}\right)!} p^{\frac{m+n}{2}} q^{\frac{n-m}{2}}, \quad (1.4)$$

где $n+m$, $n-m$ – число толчков и, следовательно, элементарных смещений (шагов) вправо и влево, соответственно.

При $p = q = \frac{1}{2}$ из (1.4)

$$P_n(m) = C_n^{\frac{n+m}{2}} \left(\frac{1}{2}\right)^n. \quad (1.5)$$

С физической точки зрения (1.5) может быть истолкована следующим образом [20]. Предположим, что в начальный момент времени $t = 0$ в точке $x = 0$ имелось достаточно большое количество частиц. Далее частицы начинают перемещаться под действием случайных толчков независимо друг от друга. Тогда (1.5) дает долю частиц, оказавшихся на расстоянии m от начала отсчета при $t \rightarrow \infty$.

При больших n для расчета вероятности $P_n(m)$, как следует из теоремы Муавра-Лапласа [20], можно использовать асимптотическую формулу

$$P_n(m) = \sqrt{\frac{2}{\pi n}} e^{-\left(\frac{m^2}{2n}\right)}. \quad (1.6)$$

Можно показать [23], используя порождающее уравнение

$$P_n(i) = \frac{P_{n+1}(i+1)}{2} + \frac{P_{n-1}(i-1)}{2}, \quad (1.7)$$

что (1.5) в непрерывном пределе действительно приводит к уравнению Фоккера-Планка.

Основные характеристики НР и соответствующие формулы для их вычисления представлены в таблице 1.1.

Таблица 1.1. Характеристики НР

| Название характеристики | Формула или числовое значение характеристики |
|------------------------------|--|
| Обозначение | $N(\mu, \sigma)$ |
| Параметры | μ, σ |
| Плотность распределения | $f_N(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \quad (1.8)$ |
| Функция распределения | $F_N(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x \exp\left[-\frac{1}{2}\left(\frac{\xi-\mu}{\sigma}\right)^2\right] d\xi \quad (1.9)$ $(x \in (-\infty, +\infty))$ |
| Среднее значение | $\mathbf{M}[x] = \mu$ |
| Дисперсия | $\mathbf{D}[x] = \sigma^2$ |
| Стандартное отклонение (СКО) | $\sqrt{\mathbf{D}[x]} = \sigma$ |
| Мода распределения | $\mathbf{Mo} = \mu$ |
| Медиана распределения | $\mathbf{Me} = \mu$ |

Из таблицы 1.1 видно, что НР, симметричное относительно точки $x = \mu$, зависит от двух параметров μ и σ , совпадающих со средним значением и СКО.

На практике оказывается удобным работать не с исходными случайными величинами x , а с соответствующими нормированными величинами z , получаемых с помощью преобразования

$$z = \frac{x - \mu}{\sigma}.$$

Распределение случайных величин z называется *стандартным нормальным распределением* (СНР). СНР с нулевым средним и единичной дисперсией обозначается $N(0,1)$. Графики функций $F(z,0,1)$, $f(z,0,1)$ представлены на рисунке 1.1.

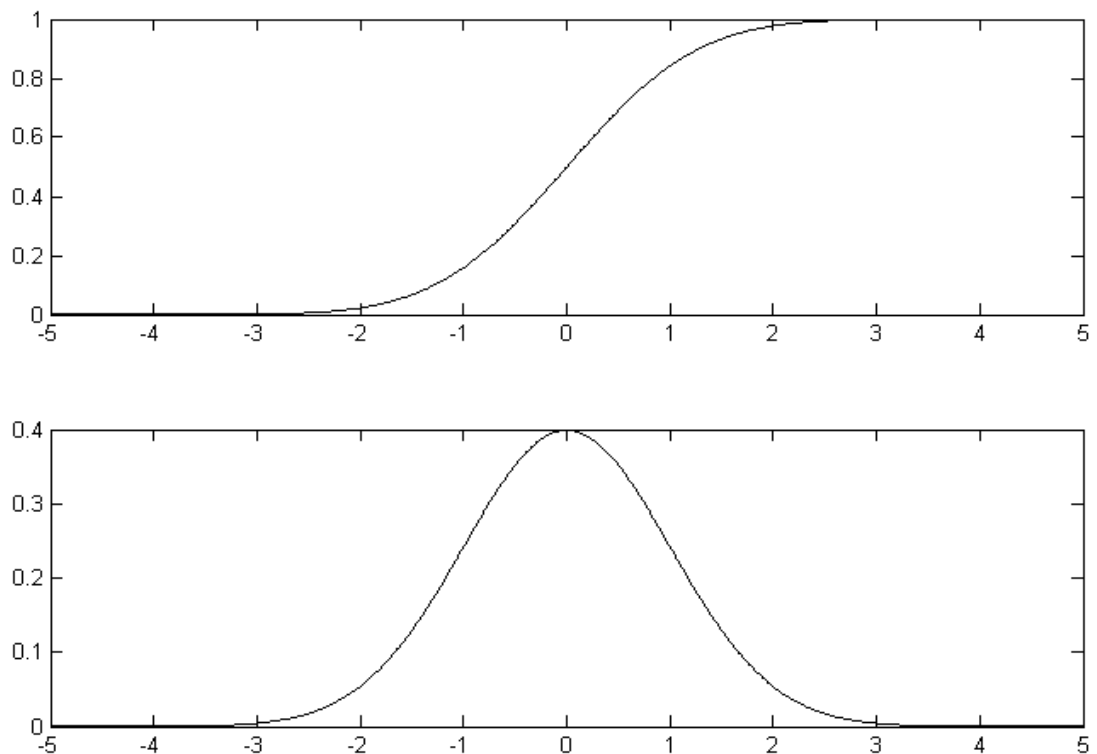


Рисунок 1.1. Нормальный закон распределения. Вверху график функции $F(z,0,1)$, внизу - $f(z,0,1)$

1.1.1.2 Усеченное нормальное распределение

Усеченное нормальное распределение (УНР) традиционно используется в задачах оценки надежности технических систем [4,21,43,50,52,53] и точности производства [9]. Напомним, что УНР $N'(\mu, \sigma)$ имеет случайная подпоследовательность $\{x'_k\}$, извлеченная из случайной последовательности $\{x_k\}$, имеющей ПР $N(\mu, \sigma)$, каждый элемент которой удовлетворяет следующим условиям:

$$x_{\min} \leq x'_k \leq x_{\max},$$

где x_{\min}, x_{\max} — точки усечения.

Основные характеристики УНР и соответствующие формулы для их вычисления представлены в таблице 1.2.

Таблица 1.2. Характеристики УНР [9]

| Название характеристики | Формула или числовое значение характеристики |
|-------------------------|--|
| Обозначение | $N'(\mu, \sigma, x_{\min}, x_{\max})$ |
| Параметры | $\mu, \sigma, x_{\min}, x_{\max}$ |
| Плотность распределения | $f_{LIM}(x; \mu, \sigma, x_{\min}, x_{\max}) = \frac{1}{\sigma} \frac{\varphi\left(\frac{x-\mu}{\sigma}\right)}{F\left(\frac{x_{\max}-\mu}{\sigma}\right) - F\left(\frac{x_{\min}-\mu}{\sigma}\right)},$ <p>где $\varphi(x)$ – плотность стандартного нормального распределения $N(0,1)$, $F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{\xi^2}{2}} d\xi$ – функция Лапласа, μ, σ – параметры исходного нормального распределения, $x_{\min} < x_{\max}$ – точки усечения</p> |
| Функция распределения | $F_{LIM}(x; \mu, \sigma, x_{\min}, x_{\max}) = \int_{a_1}^x f_{LIM}(\xi; \mu, \sigma, x_{\min}, x_{\max}) d\xi =$ $\frac{1}{\sigma} \frac{1}{F\left(\frac{x_{\max}-\mu}{\sigma}\right) - F\left(\frac{x_{\min}-\mu}{\sigma}\right)} \int_{a_1}^x \varphi\left(\frac{\xi-\mu}{\sigma}\right) d\xi =$ $\frac{F\left(\frac{x-\mu}{\sigma}\right) - F\left(\frac{a_1-\mu}{\sigma}\right)}{F\left(\frac{x_{\max}-\mu}{\sigma}\right) - F\left(\frac{x_{\min}-\mu}{\sigma}\right)}$ |
| Среднее значение | $M[x] = \mu - (\lambda_2 - \lambda_1)\sigma,$ <p>где $\lambda_1 = \frac{\varphi(\xi_1)}{F(\xi_2) - F(\xi_1)}$, $\lambda_2 = \frac{\varphi(\xi_2)}{F(\xi_2) - F(\xi_1)}$, $\xi_1 = \frac{x_{\min} - \mu}{\sigma}$, $\xi_2 = \frac{x_{\max} - \mu}{\sigma}$</p> |
| Дисперсия | $D[x] = (1 + \lambda_1 \xi_1 - \lambda_2 \xi_2 + (\lambda_1 + \lambda_2)^2) \sigma^2$ |

Из таблицы 1.2 видно, что УНР $N'(\mu, \sigma, a_1, a_2)$ относится к классу четырех параметрических распределений. Примеры ПР и ФР для различных параметров УНР представлены на рисунках П.А.1-П.А.6.

Из рисунков П.А.1-П.А.6 видно, что:

- при $x_0 \equiv 0$ и $x \in [-l; l]$ функция $f_{LIM}(x; x_0, \sigma, l)$, оказывается симметричной относительно прямой $x=0$, а центр рассеяния совпадает с математическим ожиданием, медианной и модой;
- при $x_0 \equiv 0$ и $x \in [-l; l]$ функция $F_{LAD}(x; x_0, \sigma, l)$ оказывается антисимметричной относительно прямых $x=0$, $y=0.5$;

- при $l \geq 2.5\sigma$ форма распределения практически не отличается от формы нормального распределения;
- функция $f_{LIM}(x; x_0, \sigma, l)$ на участке усечения с точностью до нормировки совпадает с функцией плотности нормального распределения $f(x; x_0, \sigma)$.

1.1.1.3 Распределение случайных последовательностей с ограниченной областью рассеяния

НР случайной величины имеет неограниченную область рассеяния. В тоже время случайные величины, анализ которых приходится проводить в естествознании, технике и экономике, имеют конечную область рассеяния. (Далее будем называть случайные числа данного типа – случайными последовательностями с ограниченной областью рассеяния (СПООР).) Априори, понятно, что ФР и ПР СПООР, будут отличаться от аналогичных величин НР (1.3).

В качестве примеров СПООР можно привести плотности углей, изменяющиеся в диапазоне от ρ_{\min} до некоторого максимального значения ρ_{\max} , [54], время безотказной работы группы однотипных приборов, изменяющегося в диапазоне от некоторого минимального значения T_{\min} до некоторого максимального T_{\max} и т.д.

Наиболее очевидной физической моделью СПООР служат песочные часы, в которых песок из «точечного» источника высыпается на горизонтальную плоскость, ограниченную непроницаемыми абсолютно упругими вертикальными стенками. При достаточном удалении стенок от источника, понятно, что они не будут оказывать влияния на формирование кучи песка, поэтому ее форма будет симметричной. В противоположном случае песчинки будут отражаться от стенок, что приведет к отличию формы кучи от «нормальной».

Анализ работ, посвященных построению моделей ФР и ПР СПООР, показывает, что существует два альтернативных подхода. Первый подход основан на описании ФР и ПР с помощью УНР [9,43]. Во втором подходе, базирующемся на работах А. Эйнштейна и М. Смолуховского по теории броуновского движения [83], модель ФР и ПР СПООР строится как распределение конечного состояния некоторого случайного процесса без последствия с ограниченной областью рассеяния [54]. Отметим, что сегодня в подавляющем большинстве учебников по теории надежности технических систем используется первый подход, как правило, без каких-либо обоснований подобного выбора [9,43]. В этой связи сравнение обоих подходов к построению математических моделей ФР и ПР СПООР с точки зрения адекватности изучаемым физическим процессам является актуальной.

ПР СПООР в рассматриваемом случае можно найти, проведя аналогию между рассматриваемой случайной величиной и одномерным броуновским

движением частицы в ограниченной области. Особенности статистических свойств данного движения были изучены М. Смолуховским [83].

Рассмотрим, следуя [20], решение обсуждаемой задачи для области рассеяния, ограниченной с правой стороны отражающей стенкой, расположенной в точке x_{\max} . Искомая вероятность нахождения броуновской частицы в точке с координатой m в рассматриваемом случае вычисляется по формуле:

$$P_n(m) \sim \frac{1}{\sqrt{2\pi n}} \left\{ e^{-\frac{m^2}{2n}} + e^{-\frac{(2x_{\max}-m)^2}{2n}} \right\}, \quad (1.10)$$

где x_{\max} – координата отражающей стенки, расположенной справа от источника, n – количество случайных толчков.

Из (1.10) видно, что вероятность нахождения броуновской частицы в произвольной точке m , может быть интерпретирована, как сложение интенсивностей двух источников (действительного и фиктивного), расположенных в точках $x_1 = 0$, $x_2 = 2x_{\max} - m$, соответственно.

Рассуждая аналогично, можно показать, что для броуновского движения с ограниченной областью рассеяния слева в точке x_{\min} вероятность нахождения броуновской частицы в точке с координатой m в данном случае вычисляется по формуле:

$$P_n(m) \sim \frac{1}{\sqrt{2\pi n}} \left\{ e^{-\frac{m^2}{2n}} + e^{-\frac{(m-2x_{\min})^2}{2n}} \right\}, \quad (1.11)$$

где x_{\min} – координата отражающей стенки, расположенной слева от источника, n – количество случайных толчков.

Из формулы (1.11) видно, что вероятность нахождения броуновской частицы в произвольной точке m , может быть интерпретирована, как сложение интенсивностей двух источников (действительного и фиктивного), расположенных в точках $x_1 = 0$, $x_2 = m - 2x_{\min}$, соответственно.

В связи с тем, что при практическом использовании модели броуновского движения в ограниченной области одним из основных оказывается вопрос о вычислении координат точек расположения фиктивных источников, рассмотрим его более подробно. Выберем систему координат с началом в середине отрезка области рассеяния (рисунок 1.2).

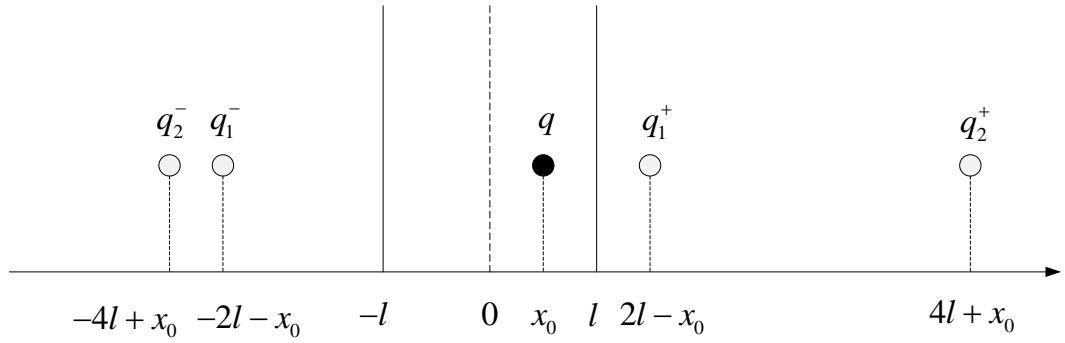


Рисунок 1.2. К вычислению координат фиктивных источников

Из рисунка 1.2 видно, что наличие двух отражающих поверхностей порождает бесконечную систему фиктивных источников. Действительно, фиктивный источник q_g^+ , полученный отражением относительно плоскости $x = +l$ и находящийся справа от нее, в свою очередь отражается относительно плоскости $x = -l$, формируя фиктивный источник $q_{g+1}^- = q_g^+$. И аналогично для фиктивного источника q_g^- . Дополнив эти соотношения правилами преобразования координат фиктивных источников q_g^\pm , получаем:

$$\begin{aligned} q_{g+1}^+ &= q_g^-, & x_{g+1}^+ - l &= -(x_g^- - l), \\ q_{g+1}^- &= q_g^+, & x_{g+1}^- + l &= -(x_g^+ + l). \end{aligned} \quad (1.12)$$

Принимая во внимание «начальные условия» $q_0^+ = q_0^- = q$, $x_0^+ = x_0^- = x_0$, из (1.12) получаем:

$$x_{2g}^\pm = \pm 4gl + x_0, \quad x_{2g+1}^\pm = \pm(4g + 2)l - x_0, \quad (1.13)$$

где $g = 0, 1, \dots$, по которым можно вычислить координаты любого из мнимых источников.

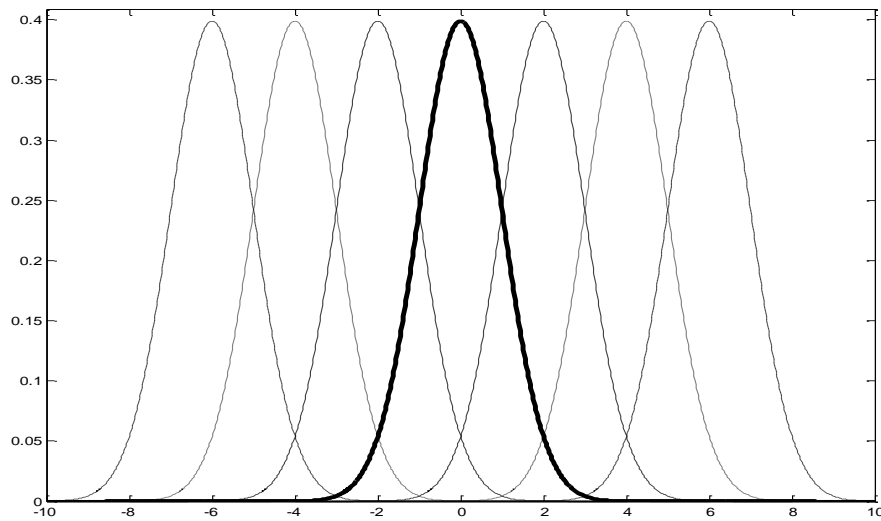


Рисунок 1.3. ПР реального и фиктивных источников. $l=1$, $x_0=0$, $x_1^+=2$, $x_1^-=-2$, $x_2^+=4$, $x_2^-=-4$, $x_3^+=6$, $x_3^-=-6$

Таким образом, используя раздельное суммирование по четным и нечетным индексам, можно записать следующее выражение для ПР СПООР:

$$f_{LAD}(x; x_0, \sigma, l) = A \left[\varphi(x; x_0, \sigma, l) + \sum_{g=0}^{\infty} \varphi_{2g+1}^{\pm}(x; x_0, \sigma, l) + \sum_{g=1}^{\infty} \varphi_{2g}^{\pm}(x; x_0, \sigma, l) \right], \quad (1.14)$$

где A – нормировочный коэффициент, определяемый из условия

$$\int_a^b f_{LAD}(\xi; x_0, \sigma, l) d\xi = 1, \quad (1.15)$$

$$\varphi(x; x_0, \sigma, l) = \exp\left[-(x - x_0)^2 / 2\sigma^2\right],$$

$$\varphi_{2g+1}^{\pm}(x; x_0, \sigma, l) = \exp\left[-(x - x_{2g+1}^{\pm})^2 / 2\sigma^2\right],$$

$$\varphi_{2g}^{\pm}(x; x_0, \sigma, l) = \exp\left[-(x - x_{2g}^{\pm})^2 / 2\sigma^2\right],$$

здесь $x_{2g+1}^{\pm}, x_{2g}^{\pm}$ вычисляются в соответствии с (1.13).

Соответственно, ФР СПООР вычисляется по формуле

$$F_{LAD}(x; x_0, \sigma, l) = \int_a^x f_{LAD}(\xi; x_0, \sigma, l) d\xi. \quad (1.16)$$

Необходимо отметить, что плотность распределения нормальной случайной величины (1.14), вообще говоря, формируется бесконечной системой фиктивных источников. Однако для расчетов на ЭВМ достаточно некоторого конечного числа фиктивных источников. На рисунке 1.4 представлена зависимость

$$\varepsilon(N, \mu, \sigma, x_{\min}, x_{\max}) = \ln \left[\max \left(|f_{LAD}(x; \mu, \sigma, x_{\min}, x_{\max}, N) - f_{LAD}(x; \mu, \sigma, x_{\min}, x_{\max}, 100)| \right) \right],$$

вычисленная для следующих значений параметров:
 $\mu = 0, \sigma = 0, x_{\min} = -5, x_{\max} = 5$.

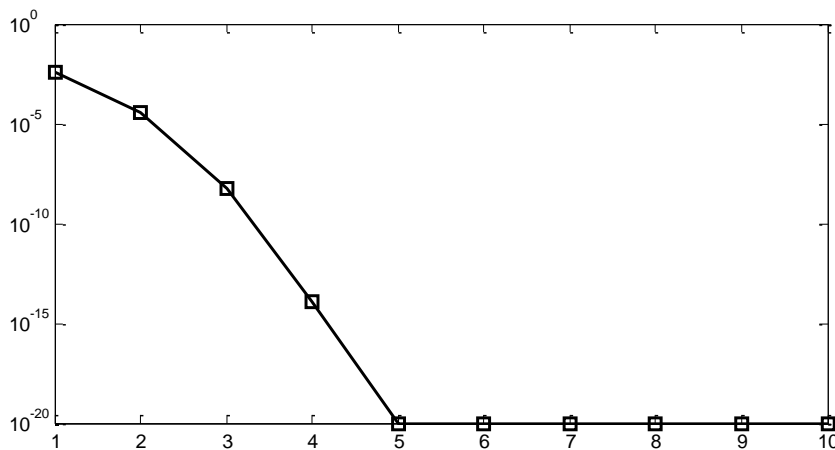


Рисунок 1.4. График зависимости $\varepsilon(N, 0, 10, -5, 5)$

Как видно из рисунка 1.4, достаточная точность вычисления плотности вероятности в соответствие с (1.14) достигается при $N=5$ пар мнимых источников (соответственно, число мнимых источников – $2N=10$), а потому указанную величину можно не оценивать с помощью ГА.

Характеристики рассматриваемого распределения и соответствующие формулы для их вычисления представлены в таблице 1.3.

Таблица 1.3. Характеристики ПР СПООР

| Название характеристики | Формула или числовое значение характеристики |
|-------------------------|---|
| Обозначение | $LAD(x_0, \sigma, x_{\min}, x_{\max})$ |
| Параметры | $\mu, \sigma, x_{\min}, x_{\max}$ |
| Плотность распределения | $f_{LAD}(x; x_0, \sigma, x_{\min}, x_{\max}) = A \left[\varphi(x; x_0, \sigma, x_{\min}, x_{\max}) + \sum_{g=0}^{\infty} \varphi_{2g+1}^{\pm}(x; x_0, \sigma, x_{\min}, x_{\max}) + \sum_{g=1}^{\infty} \varphi_{2g}^{\pm}(x; x_0, \sigma, x_{\min}, x_{\max}) \right],$ <p>где A – нормировочный коэффициент, определяемый из условия</p> $\int_{x_{\min}}^{x_{\max}} f_{LAD}(\xi; x_0, \sigma, x_{\min}, x_{\max}) d\xi = 1,$ $\varphi(x; x_0, \sigma, x_{\min}, x_{\max}) = \exp\left[-(x - \mu)^2 / 2\sigma^2\right],$ $\varphi_{2g+1}^{\pm}(x; x_0, \sigma, x_{\min}, x_{\max}) = \exp\left[-(x - x_{2g+1}^{\pm})^2 / 2\sigma^2\right],$ $\varphi_{2g}^{\pm}(x; x_0, \sigma, x_{\min}, x_{\max}) = \exp\left[-(x - x_{2g}^{\pm})^2 / 2\sigma^2\right],$ $x_{2g}^{\pm} = \pm 4g(x_{\max} - x_{\min}) + \mu, \quad x_{2g+1}^{\pm} = \pm(4g + 2)(x_{\max} - x_{\min}) - \mu,$ $x_{\min} < x_{\max} \text{ – границы интервала рассеяния}$ |
| Функция распределения | $F_{LAD}(x; x_0, \sigma, x_{\min}, x_{\max}) = \int_{x_{\min}}^x f_{LAD}(\xi; x_0, \sigma, x_{\min}, x_{\max}) d\xi.$ |

Отметим, что, формально, ФР, вычисляемая в соответствие с (1.16), является 3х-параметрическим (оно зависит от следующих параметров: x_0 – положение центра рассеяния, σ – СКО при отсутствии ограничения, l – размаха области рассеяния). Однако если принять во внимание, что

$$l = x_{\max} - x_{\min},$$

где x_{\min}, x_{\max} – координаты, левой и правой границ области рассеяния, знание значений которых оказывается весьма важным в практических приложениях, распределение (1.16) следует отнести к классу 4-х-параметрических распределений.

Анализ свойств функций $f_{LAD}(x; 0, \sigma, l)$, $F_{LAD}(x; 0, \sigma, l)$, представленных на рисунках 1.5, 1.6, показывает, что:

- при $x_0 \equiv 0$ функция $f_{LAD}(x; 0, \sigma, l)$, оказывается симметричной относительно прямой $x=0$, а центр рассеяния совпадает с математическим ожиданием, медианной и модой;
- при $x_0 \equiv 0$ функция $F_{LAD}(x; 0, \sigma, l)$ оказывается антисимметричной относительно прямых $x=0$, $y=0.5$;
- при $l \geq 2.5\sigma$ влияние фиктивных источников уменьшается и рассеяние приближается к нормальному закону;
- по мере уменьшения размера области рассеяния влияние фиктивных источников возрастает и при $l \leq 0.7\sigma$ распределение выглядит как равномерное.

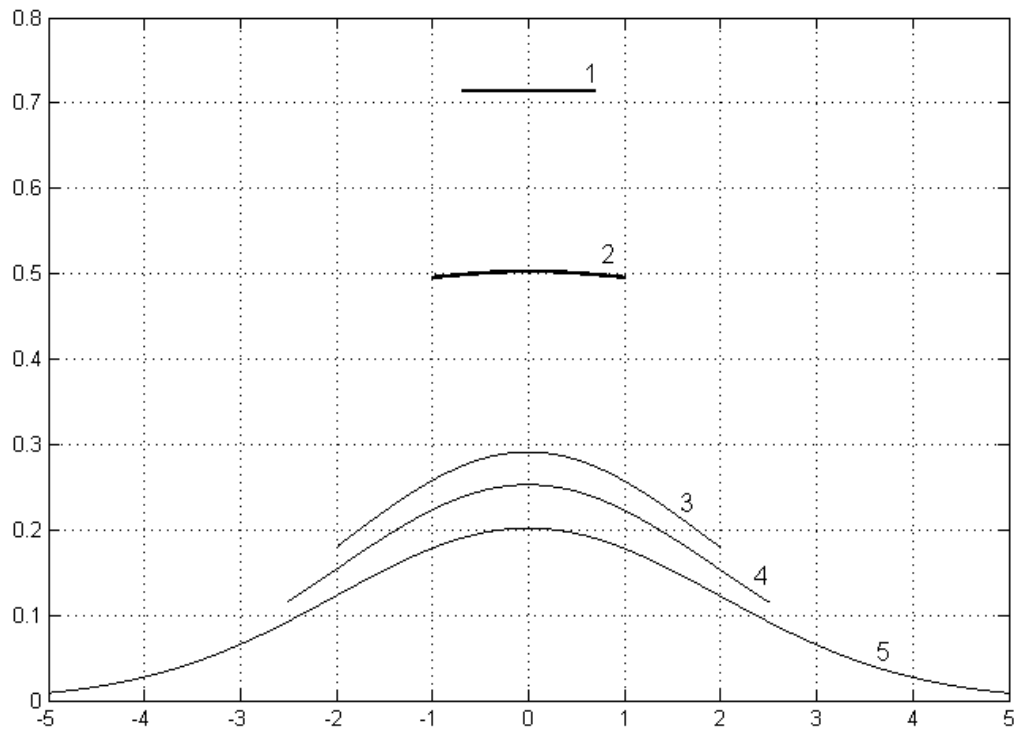


Рисунок 1.5. СПООР: график функции $f_{LAD}(x; x_0, \sigma, l)$:

- 1 – $x_0 = 0, 2l = 0.7\sigma, x \in [-l, l]$; 2 – $x_0 = 0, 2l = 1.0\sigma, x \in [-l, l]$; 3 – $x_0 = 0, 2l = 2.0\sigma, x \in [-l, l]$;
 4 – $x_0 = 0, 2l = 2.5\sigma, x \in [-l, l]$; 5 – $x_0 = 0, 2l = 5.0\sigma, x \in [-l, l]$

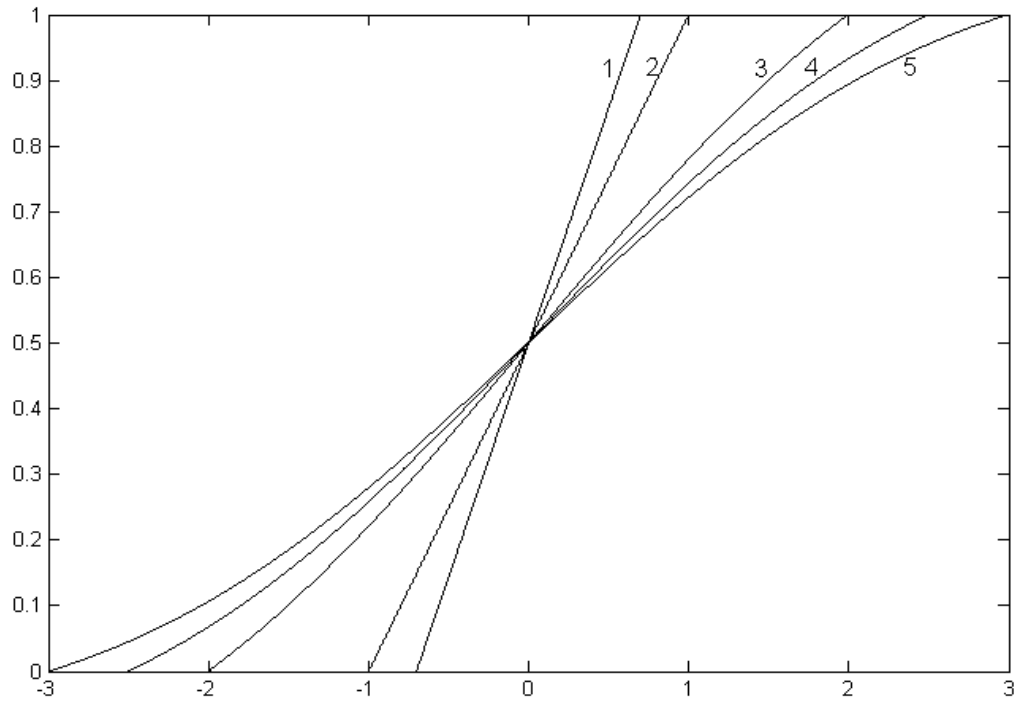


Рисунок 1.6. СПООР: график функции $F_{LAD}(x; x_0, \sigma, l)$:

1 – $x_0 = 0, 2l = 0.7\sigma, x \in [-l, l]$; 2 – $x_0 = 0, 2l = 1.0\sigma, x \in [-l, l]$; 3 – $x_0 = 0, 2l = 2.0\sigma, x \in [-l, l]$;

4 – $x_0 = 0, 2l = 2.5\sigma, x \in [-l, l]$; 5 – $x_0 = 0, 2l = 5.0\sigma, x \in [-l, l]$

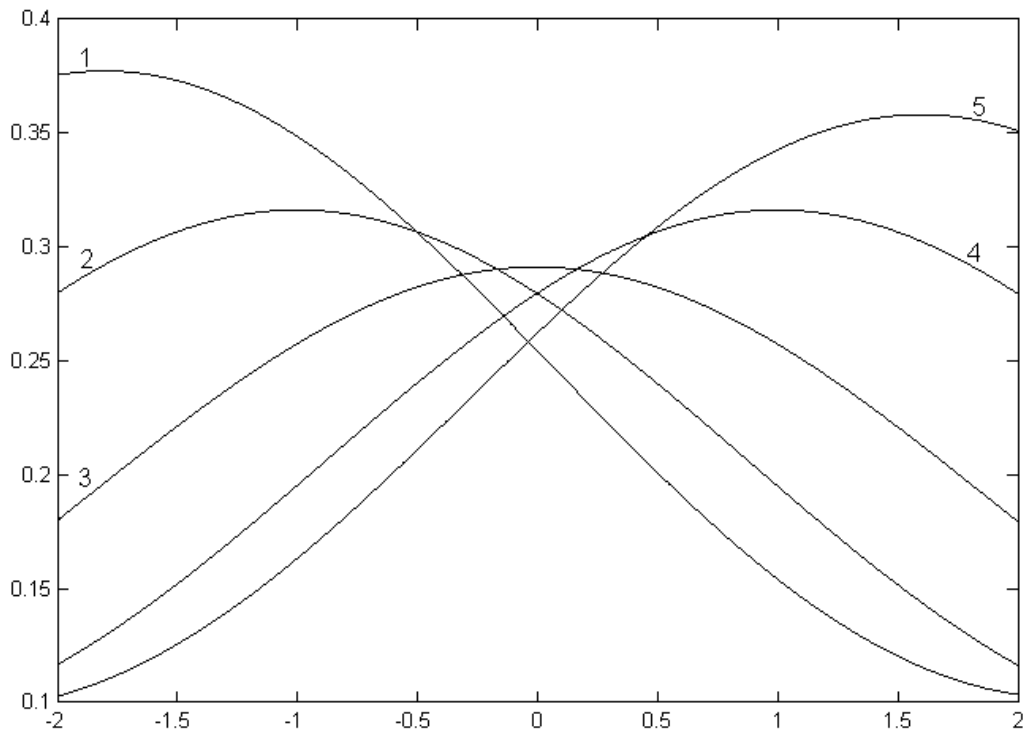


Рисунок 1.7. СПООР: график функции $f_{LAD}(x; x_0, \sigma, l/\sigma)$:

1 – $x_0 = -0.9\sigma, 2l = 2\sigma$; 2 – $x_0 = -0.5\sigma, 2l = 2\sigma$; 3 – $x_0 = 0, 2l = 2.0\sigma$; 4 – $x_0 = 0.5\sigma, 2l = 2.0\sigma$;

5 – $x_0 = 0.9\sigma, 2l = 2.0\sigma$

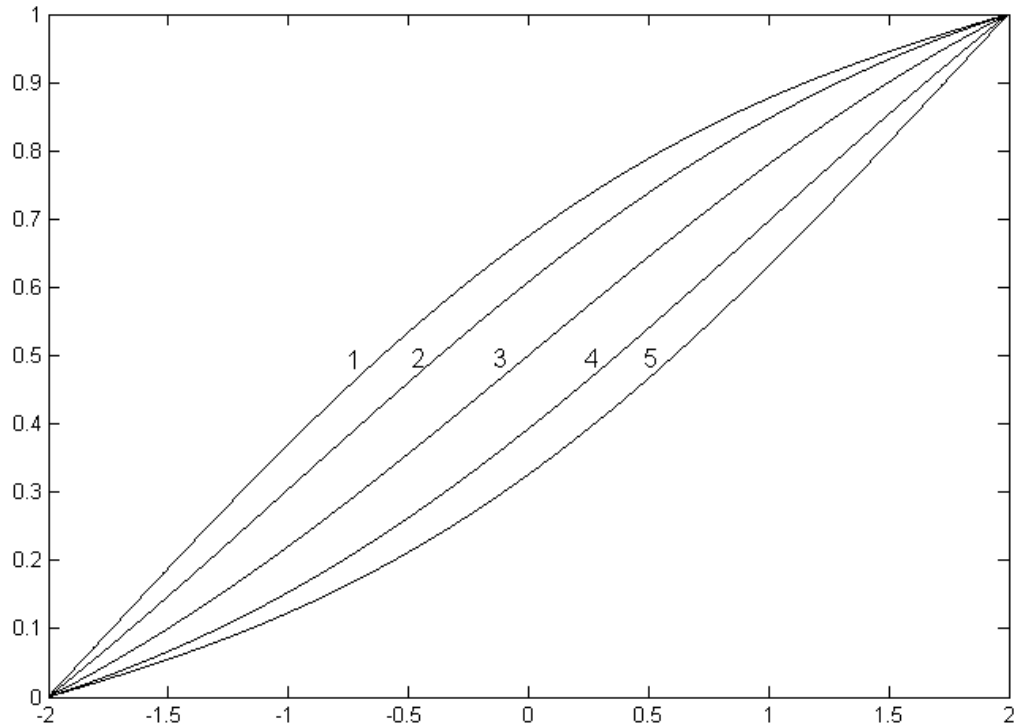


Рисунок 1.8. СПООР: график функции $F_{LAD}(x; x_0, \sigma, l)$:

1 – $x_0 = -0.9\sigma, 2l = 2\sigma$; 2 – $x_0 = -0.5\sigma, 2l = 2\sigma$; 3 – $x_0 = 0, 2l = 2.0\sigma$;

4 – $x_0 = 0.5\sigma, 2l = 2.0\sigma$; 5 – $x_0 = 0.9\sigma, 2l = 2.0\sigma$

При $l \in [0.7, 2.5]\sigma$ распределение (1.14), (1.16), порождает семейство распределений, изменяющихся по своей форме от НР до равномерного распределения.

Анализ свойств функций $f_{LAD}(x; x_0, \sigma, 2\sigma)$, $F_{LAD}(x; x_0, \sigma, 2\sigma)$, представленных на рисунках 1.7, 1.8, позволяет сделать следующие выводы:

- при $x_0 \neq 0$ плотность вероятности $f_{LAD}(x; 0, \sigma, l)$, вычисляемая в соответствие с (1.14), и функция распределения $F_{LAD}(x; 0, \sigma, l)$ вычисляемая в соответствие с (1.16), оказываются асимметричными;
- коэффициент асимметрии изменяется в интервале $[-1, 1]$, он равен: -1 при $x_0 = -1$, 0 при $x_0 = 0$, 1 при $x_0 = 1$.

Квантиль функции нормального распределения с ограниченной областью рассеяния есть функция, зависящая от доверительной вероятности:

$$x_p^{(LAD)} = \arg(F_{LAD}(x; x_0, \sigma, l) = p) = Q(p; x_0, \sigma, l)$$

Зависимости $Q(\alpha, x_0, \sigma, l)$ при заданных значениях α, x_0, σ от длины области рассеяния l представлены на рисунках 1.9–1.11.

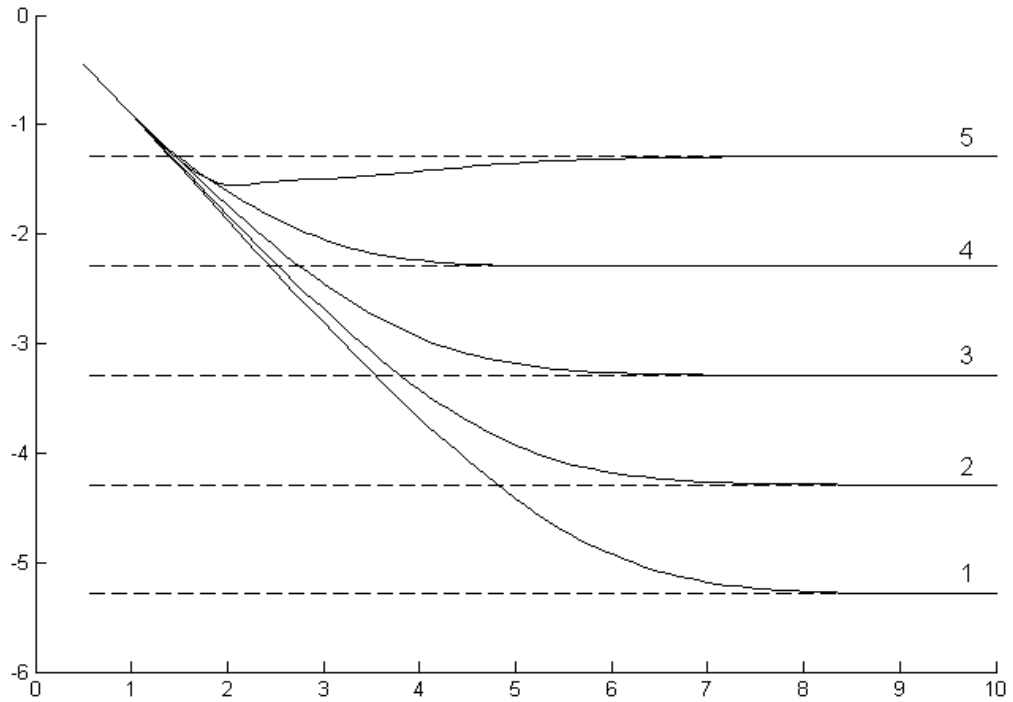


Рисунок 1.9. СПООР: график функции $Q(0.05; x_0, \sigma, l/\sigma)$, $l \in [0.5, 10]\sigma$

1 – $x_0 = -\sigma$, 2 – $x_0 = -0.5\sigma$, 3 – $x_0 = 0$, 4 – $x_0 = 0.5\sigma$, 5 – $x_0 = \sigma$

(пунктирные линии – соответствующие квантили равномерного распределения с неограниченной областью рассеяния)

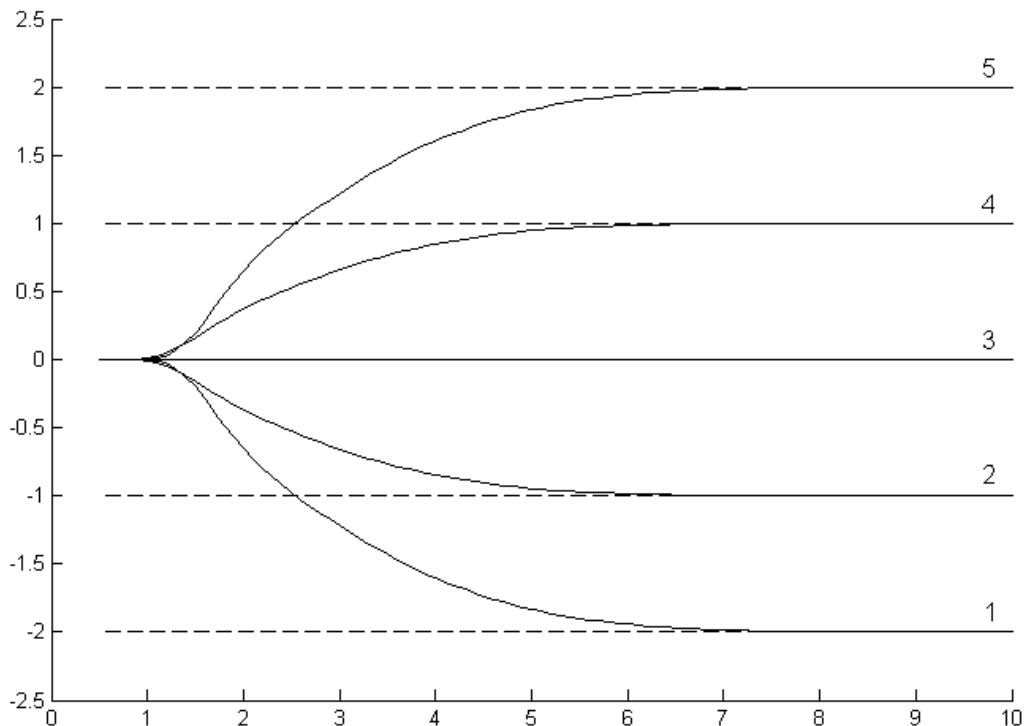


Рисунок 1.10. СПООР: график функции $Q(0.5; x_0, \sigma, l/\sigma)$, $l \in [0.5, 10]\sigma$

1 – $x_0 = -\sigma$, 2 – $x_0 = -0.5\sigma$, 3 – $x_0 = 0$, 4 – $x_0 = 0.5\sigma$, 5 – $x_0 = \sigma$

(пунктирные линии – соответствующие квантили равномерного распределения с неограниченной областью рассеяния)

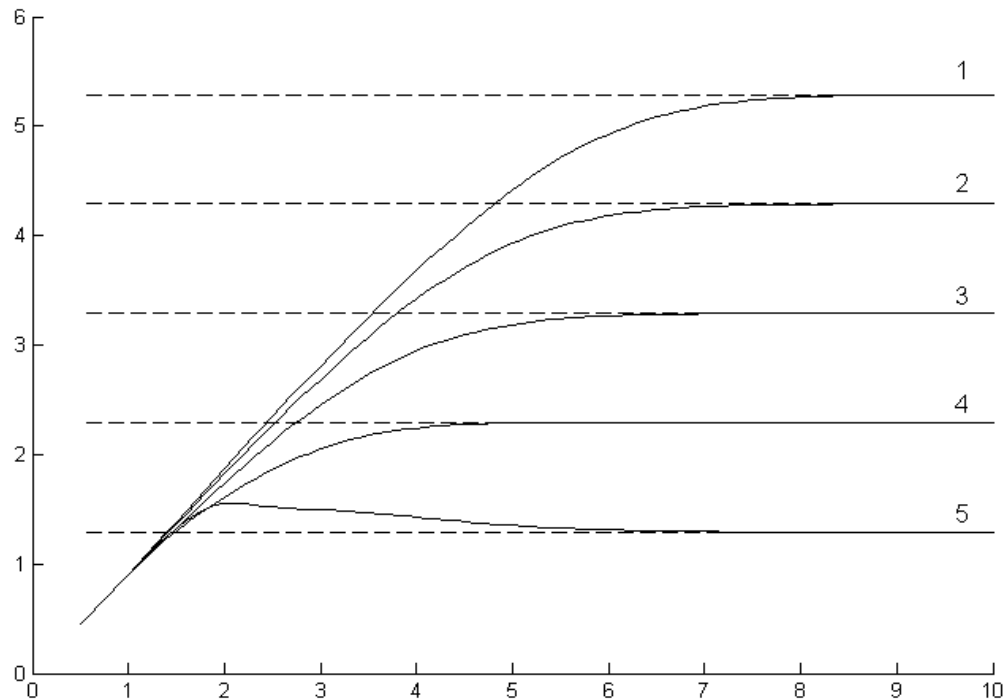


Рисунок 1.11. СПООР: график функции $Q(0.95; x_0, \sigma, l/\sigma)$, $l \in [0.5, 10]\sigma$

1 – $x_0 = -\sigma$, 2 – $x_0 = -0.5\sigma$, 3 – $x_0 = 0$, 4 – $x_0 = 0.5\sigma$, 5 – $x_0 = \sigma$

(пунктирные линии – соответствующие квантили равномерного распределения с неограниченной областью рассеяния)

Из рисунков 1.9–1.11 видно, что значения ФР СПООР с ограниченной областью рассеяния оказываются, существенно зависящими от параметров распределения. Например, при нахождении источника в центре области рассеяния ($x_0 = 0$), значения $x_{0.05}^{(LAD)}$, $x_{0.95}^{(LAD)}$ совпадают с соответствующими значениями квантилей нормального закона распределения при размере области рассеяния $l \geq 7\sigma$. Данный результат следует принимать во внимание, при анализе функций распределения случайных данных, имеющих ограниченную область рассеяния.

Примеры использования ФР и ПР СПООР при анализе эмпирической информации (распределения содержания углерода в угле, распределения прочности угля, распределения скорости проходки шурупов на угольных разрезах) приведены в [54].

1.1.1.4 Отличия между усеченным нормальным распределением и распределением случайной последовательности с ограниченной областью блуждания

Из физических соображений понятно, что ПР и ФР выборок случайных чисел, область значений которых является конечной, следует использовать модели (1.14), (1.16). (Для нахождения их численного значения может быть, например, использован метод максимального правдоподобия, приводящий к поиску решения сложной системы нелинейных уравнений).

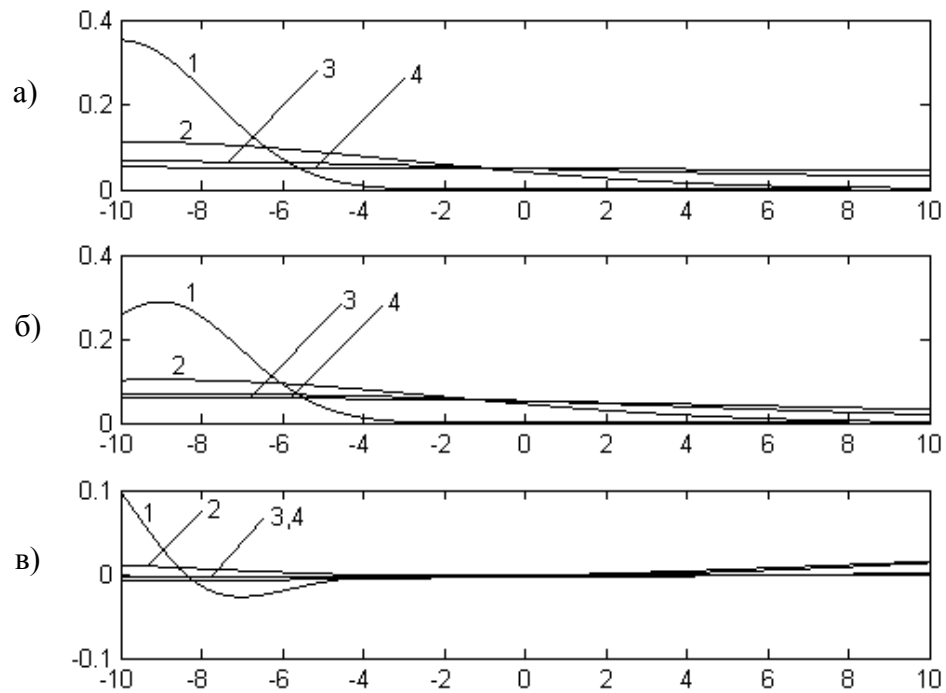


Рисунок 1.12. Сравнение УНР и СПООР: а) – график ПР СПООР, описываемой (1.14); б) – график ПР УНР; в) – разность ПР СПООР и УНР (параметры функций: $x_{\min} = -10$, $x_{\max} = 10$, $\mu = -9$, $1 - \sigma = 2$; $2 - \sigma = 7$; $3 - \sigma = 12$; $4 - \sigma = 17$)

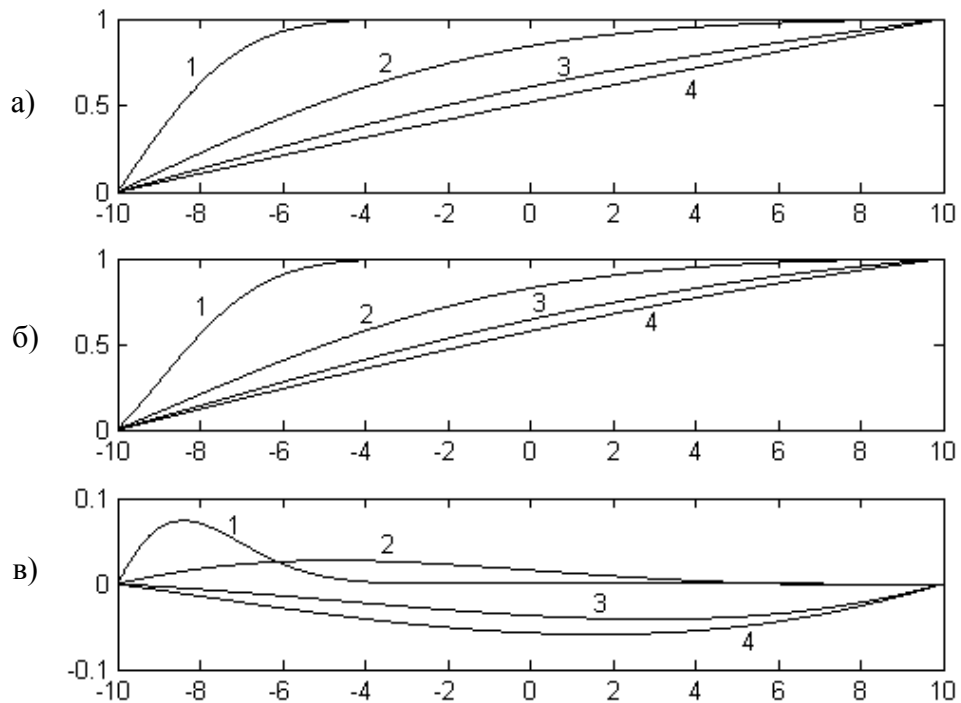


Рисунок 1.13. Сравнение УНР и СПООР: а) – график ФР, описываемой (1.16); б) – ФР УНР усеченное нормальное распределение; в) – разность ПР СПООР и УНР (параметры функций: $x_{\min} = -10$, $x_{\max} = 10$, $\mu = -9$, $1 - \sigma = 2$; $2 - \sigma = 7$; $3 - \sigma = 12$; $4 - \sigma = 17$)

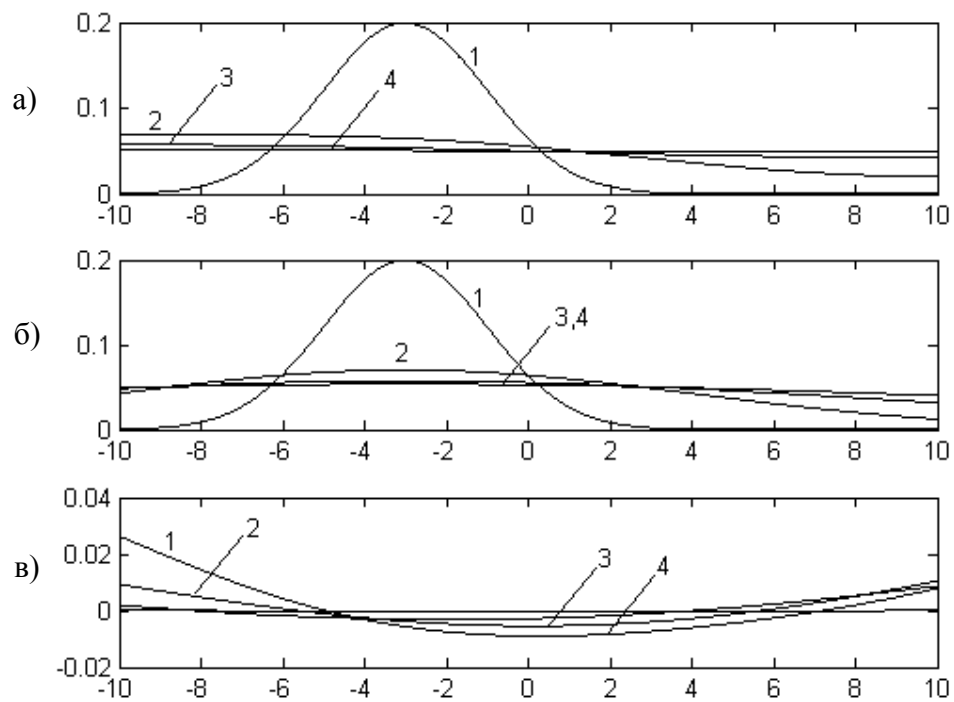


Рисунок 1.14. Сравнение УНР и СПООР: а) – график ПР СПООР, описываемой (1.14); б) – график ПР УНР; в) – разность ПР СПООР и УНР (параметры функций: $x_{\min} = -10$, $x_{\max} = 10$, $\mu = -3$, $1 - \sigma = 2$; $2 - \sigma = 7$; $3 - \sigma = 12$; $4 - \sigma = 17$)

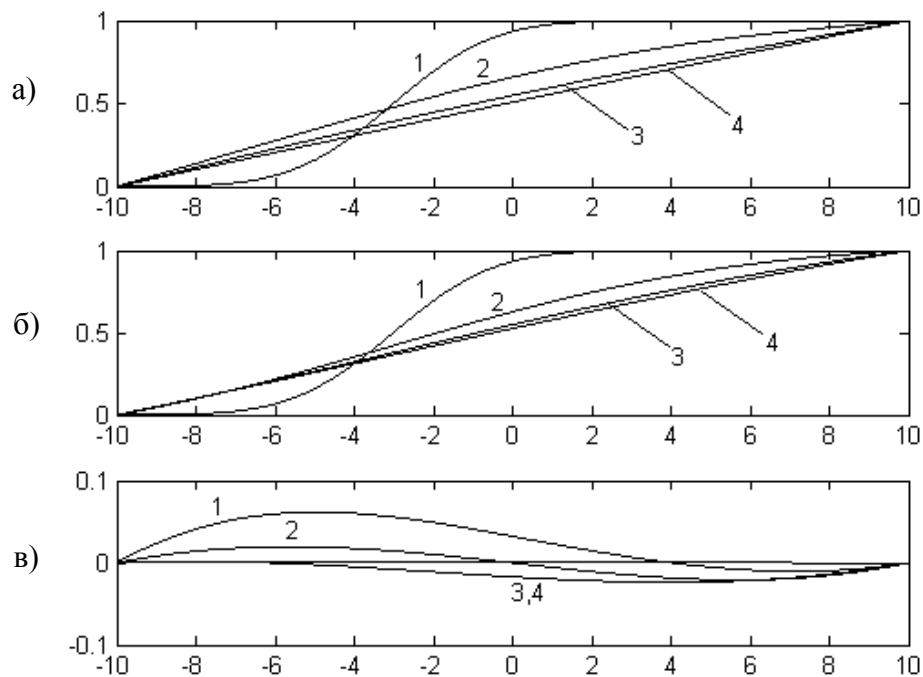


Рисунок 1.15. Сравнение УНР и СПООР: а) – график ФР, описываемой (1.16); б) – ФР УНР усеченное нормальное распределение; в) – разность ПР СПООР и УНР (параметры функций: $x_{\min} = -10$, $x_{\max} = 10$, $\mu = -3$, $1 - \sigma = 2$; $2 - \sigma = 7$; $3 - \sigma = 12$; $4 - \sigma = 17$)

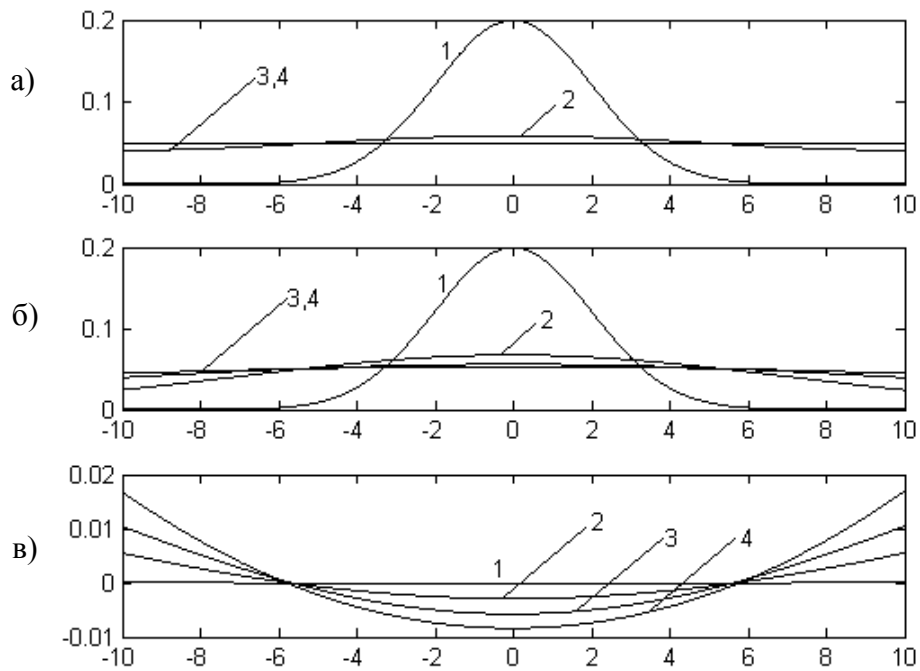


Рисунок 1.16. Сравнение УНР и СПООР: а) – график ПР СПООР, описываемой (1.14); б) – график ПР УНР; в) – разность ПР СПООР и УНР (параметры функций: $x_{\min} = -10$, $x_{\max} = 10$, $\mu = 0$, $1 - \sigma = 2$; $2 - \sigma = 7$; $3 - \sigma = 12$; $4 - \sigma = 17$)

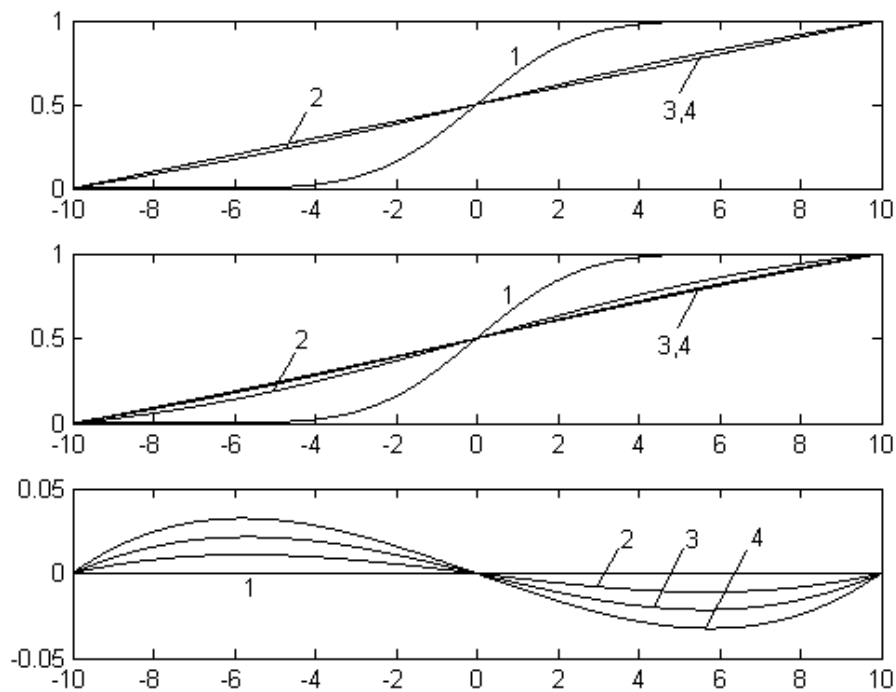


Рисунок 1.17. Сравнение УНР и СПООР: а) – график ФР, описываемой (1.16); б) – ФР УНР усеченное нормальное распределение; в) – разность ПР СПООР и УНР (параметры функций: $x_{\min} = -10$, $x_{\max} = 10$, $\mu = 0$, $1 - \sigma = 2$; $2 - \sigma = 7$; $3 - \sigma = 12$; $4 - \sigma = 17$)

Из рисунков 1.12–1.17 видно, что наименьшие отличия между ФР и ПР СПООР и УНР оказываются при расположении точечного источника в точке

$x = 0$, соответственно, $\mu = 0$. При этом увеличение значения σ приводит к увеличению данных отличий, максимальные значения которых могут достигать 1,8% и 4,8% для ПР и ФР, соответственно. Смещение положения источника от центра области рассеяния также приводит к увеличению отличий между анализируемыми зависимостями: максимальные значения разностей между ПР и ФР изучаемых случайных последовательностей для случая $\sigma = 2, \mu = -9$ составляют 10% и 9,8%, соответственно; для случая $\sigma = 2, \mu = -3$ – 3% и 7,5%, соответственно. Таким образом, в связи с тем, что при анализе случайной последовательности, у которой априори неизвестны положение источника и параметр σ , использование УНР может приводить к возникновению значимых погрешностей при оценивании ПР и ФР. Результаты исследования, приведенные в этом разделе, опубликованы в [67].

1.1.2 Анализ особенностей двумодальных распределений случайных последовательностей с ограниченной областью рассеяния

В связи с тем, что на практике, например, при оценке прочностной надежности изделий [74] или анализе суточной выработки экскаваторов на горных работах [54] требуется получение оценок двумодальных ФР и ПР, поэтому разработка методов оценки их параметров является актуальной задачей.

Отметим, что двумодальные распределения достаточно часто встречаются в природе. Например, именно таковыми являются ПР типов галактик по цвету в Слоановском цифровом обзоре неба [123], а также эффекта приспособленности [98,104].

1.1.2.1 Двумодальное распределение, представляющее линейную комбинацию НР

Двумодальные ФР и ПР представляет собой линейные комбинации ПР НР $f_{N1}(x, \alpha_1, \sigma_1)$, $f_{N2}(x, \alpha_2, \sigma_2)$, вычисляемые в соответствии с (1.8), и ФР НР $F_{N1}(x)$, $F_{N2}(x)$, вычисляемые в соответствии с (1.9):

$$f(x, \alpha_1, \sigma_1, \alpha_2, \sigma_2) = f_{N1}(x, \alpha_1, \sigma_1) \cdot \alpha + f_{N2}(x, \alpha_2, \sigma_2) \cdot (1 - \alpha), \quad (1.17)$$

где α – коэффициент смешивания, $0 < \alpha < 1$,

$$F(x, \alpha_1, \sigma_1, \alpha_2, \sigma_2) = F_{N1}(x, \alpha_1, \sigma_1) \cdot \alpha + F_{N2}(x, \alpha_2, \sigma_2) \cdot (1 - \alpha). \quad (1.18)$$

Пример двумодальных ПР и ФР, полученной в результате сложения двух одномодальных НР, представлен на рисунках 1.18, 1.19.

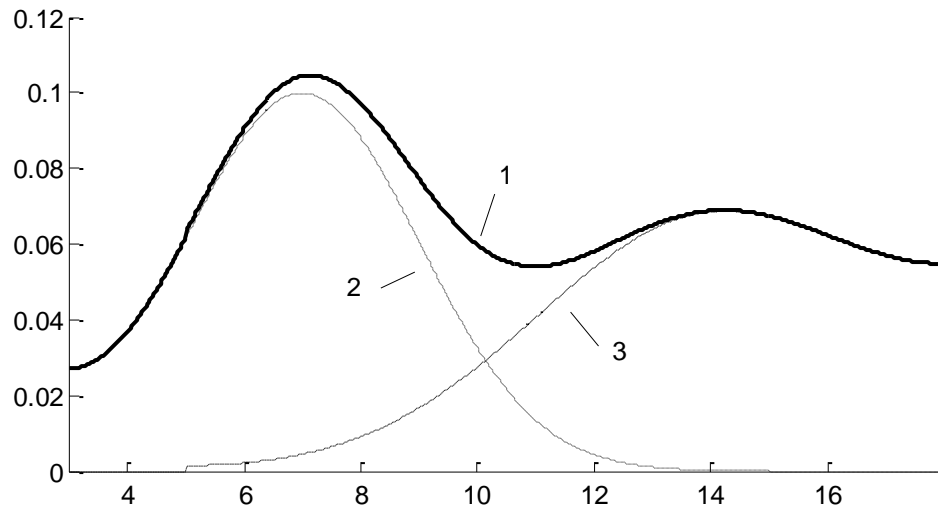


Рисунок 1.18. Пример двумодальной ПР: 1 – график двумодальной ПР, вычисленной в соответствии с (1.17) распределения случайной величины $f_{N1}(x, \alpha_1, \sigma_1)$, 2 – график функции $f_{N1}(x, \alpha_1, \sigma_1)$, 3 – график функции $f_{N2}(x, \alpha_2, \sigma_2)$, параметры распределения: $\alpha_1 = 7, \sigma_1 = 2, \alpha_2 = 14, \sigma_2 = 3, \alpha = 0.5$

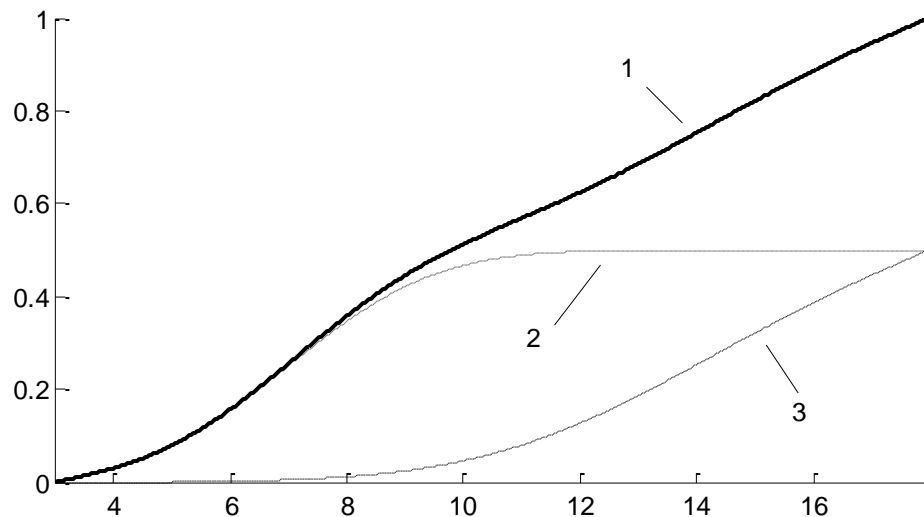


Рисунок 1.19. Пример двумодальной ФР: 1 – график ФР, вычисленной в соответствии с (1.18), 2 – график функции $F_1(x, \mu_1, \sigma_1)$, 3 – график функции $F_2(x, \mu_2, \sigma_2)$, $\alpha_1 = 7, \sigma_1 = 2, \alpha_2 = 14, \sigma_2 = 3, \alpha = 0.5$

Вообще говоря, возможны различные комбинации параметров исходных распределений, при которых суммарное распределение не будет иметь ярко выраженных экстремумов (рисунки 1.20, 1.21).

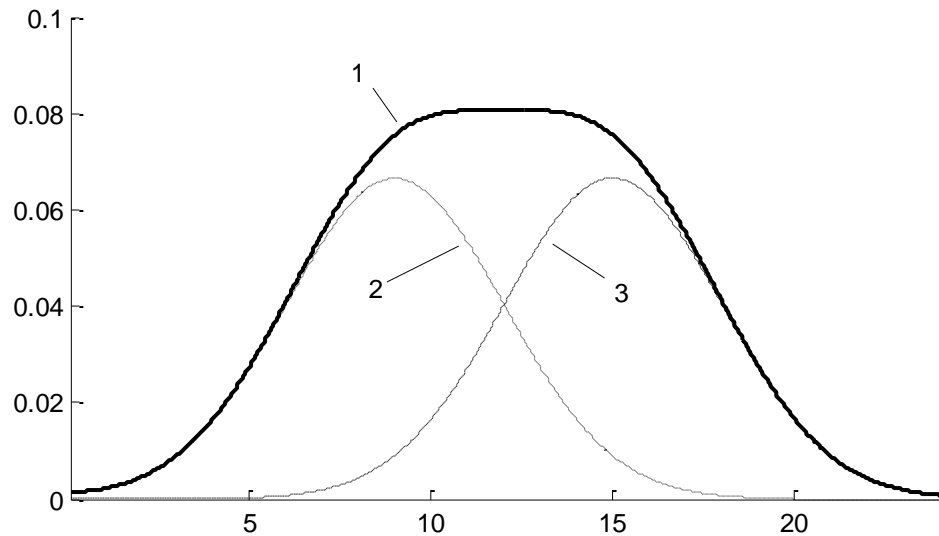


Рисунок 1.20. Пример двумодальной ПР: 1 – график двумодальной ПР, вычисленной в соответствии с (1.17) распределения случайной величины $f_{N1}(x, \alpha_1, \sigma_1)$, 2 – график функции $f_{N1}(x, \alpha_1, \sigma_1)$, 3 – график функции $f_{N2}(x, \alpha_2, \sigma_2)$, параметры распределения: $\mu_1 = 9, \sigma_1 = 3, \mu_2 = 15, \sigma_2 = 3, \alpha = 0.5$

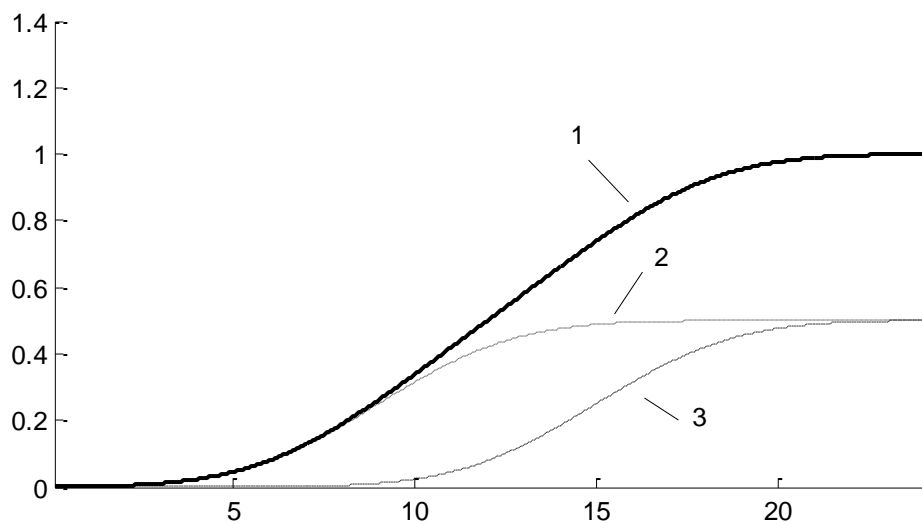


Рисунок 1.21. Пример двумодальной ПР: 1 – график двумодальной ПР, вычисленной в соответствии с (1.18), 2 – график функции $f_{N1}(x, \alpha_1, \sigma_1)$, 3 – график функции $f_{N2}(x, \alpha_2, \sigma_2)$, параметры распределения: $\mu_1 = 9, \sigma_1 = 3, \mu_2 = 15, \sigma_2 = 3, \alpha = 0.5$

Подобными ФР и ПР обладают, например, распределение роста людей. ПР роста мужчин и женщин по отдельности представляют собой одномодальные НР с разными средними значениями и дисперсиями. Авторы исследования, описанного в [107,126], утверждают, что гистограмма смешанной выборки, состоящей из мужчин и женщин, оказывается двумодальной. Однако исследование проводилось над выборкой малого объема, состоящей из данных о студентах одного университета. В [113] проведено более детальное исследование над данными гораздо большего объема, полученными из Национального центра статистики здоровья США, и

показано, что суммарная плотность распределения не обладает ярко выраженными модами и имеет вид аналогичный зависимости, представленной на рисунке 1.20.

Существуют аналитические критерии, позволяющие проверить гипотезу о том, что функция распределения случайной последовательности является двумодальной.

Критерий №1 был предложен в [103]:

Пусть $f_{N1}(x, \alpha_1, \sigma_1)$, и $f_{N2}(x, \alpha_2, \sigma_2)$, – ПР НР с математическими ожиданиями μ_1 и μ_2 и СКО $\sigma_1 = \sigma_2 = \sigma$, соответственно. Тогда распределение

$$f(x, \alpha_1, \sigma_1, \alpha_2, \sigma_2) = 0.5 \cdot f_1(x, \alpha_1, \sigma_1) + 0.5 \cdot f_2(x, \alpha_2, \sigma_2).$$

будет одномодальным тогда и только тогда, когда выполняется условие

$$\frac{|\mu_1 - \mu_2|}{2\sigma} \leq 1. \quad (1.19)$$

Критерий № 2, описанный в [87], является более общим и не зависит от соотношения дисперсий, но жестко привязан к коэффициенту смешивания. Он имеет следующую формулировку:

Пусть $f_{N1}(x, \alpha_1, \sigma_1)$, и $f_{N2}(x, \alpha_2, \sigma_2)$, – функции плотности вероятностей нормальных распределений с математическими ожиданиями и СКО μ_1, σ_1 и μ_2, σ_2 соответственно. Тогда распределение

$$f(x, \alpha_1, \sigma_1, \alpha_2, \sigma_2) = 0.5 f_{N1}(x, \alpha_1, \sigma_1) \cdot \alpha + 0.5 f_{N2}(x, \alpha_2, \sigma_2),$$

является двумодальным, если выполняется условие:

$$\frac{|\mu_1 - \mu_2|}{\sqrt{2(\sigma_1^2 + \sigma_2^2)}} > 1. \quad (1.20)$$

Критерий №3 [105], не зависящий от коэффициента смешивания, но накладывающий ограничения на соотношение дисперсий, (является расширением критерия №1) имеет следующую формулировку:

Пусть $f_{N1}(x, \alpha_1, \sigma_1)$ и $f_{N2}(x, \alpha_2, \sigma_2)$ – ПР НР с математическими ожиданиями μ_1 и μ_2 и СКО $\sigma_1 = \sigma_2 = \sigma$ соответственно. Обозначим

$$d = \frac{|\mu_1 - \mu_2|}{2\sigma}. \quad (1.21)$$

Тогда ПР (1.17) будет одномодальным тогда и только тогда, когда выполняется одно из следующих условий:

$$d \leq 1, \quad (1.22)$$

$$d > 1 \text{ и } |\log(\alpha) - \log(1 - \alpha)| \geq 2 \log(d - \sqrt{d^2 - 1}) + 2d\sqrt{d^2 - 1}. \quad (1.23)$$

Критерий №4 [113] применим к любым коэффициентам смешивания и соотношениям дисперсий:

Пусть $f_1(x)$ и $f_2(x)$ - ПР НР с математическими ожиданиями и СКО μ_1, σ_1 и μ_2, σ_2 соответственно. Обозначим

$$r = \frac{\sigma_1^2}{\sigma_2^2}, \quad (1.24)$$

$$S(r) = \frac{\sqrt{-2 + 3r + 3r^2 - 2r^3 + 2(1 - r + r^2)^{3/2}}}{\sqrt{r}(1 + \sqrt{r})}. \quad (1.25)$$

Тогда распределение (1.17) является одномодальным, если выполняется условие:

$$\frac{|\mu_1 - \mu_2|}{S(r)(\sigma_1 + \sigma_2)} \leq 1. \quad (1.26)$$

Основные характеристики двумодального распределения и соответствующие формулы для их вычисления представлены в таблице 1.4 [111].

Таблица 1.4. Характеристики двумодального распределения

| Название характеристики | Формула или числовое значение характеристики |
|-------------------------|---|
| Обозначение | $BI(x; \mu_1, \sigma_1, \mu_2, \sigma_2, \alpha)$ |
| Параметры | $\mu_1, \sigma_1, \mu_2, \sigma_2, \alpha$ |
| Плотность распределения | $f_{BI}(x) = f(x, \mu_1, \sigma_1) \cdot \alpha + f(x, \mu_2, \sigma_2) \cdot (1 - \alpha)$ <p>где $f(x)$ – плотность одномодального нормального распределения μ_1, σ_1 – параметры первой составляющей распределения, μ_2, σ_2 – параметры второй составляющей распределения, α – коэффициент смешивания</p> |
| Функция распределения | $F_{BI}(x) = F(x, \mu_1, \sigma_1) \cdot \alpha + F(x, \mu_2, \sigma_2) \cdot (1 - \alpha)$ |
| Среднее значение | $M[x] = \mu = \alpha \cdot \mu_1 + (1 - \alpha) \cdot \mu_2$ |
| Дисперсия | $D[x] = \sigma^2 = \alpha \cdot (\sigma_1^2 + (\mu_1 - \mu)^2) + (1 - \alpha) \cdot (\sigma_2^2 + (\mu_2 - \mu)^2)$ |

Отметим, что классическое двумодальное распределение является 5-ти параметрическим. Основные характеристики распределения – среднее значение и дисперсия – могут быть вычислены аналитически из характеристик составляющих распределений.

1.1.2.2 Двумодальное распределение с ограниченной областью рассеяния

В данной работе рассмотрено двумодальное распределение ограниченной случайной последовательности (ДРОСП), каждая мода ПР которого является ПР СПООР.

Функция изучаемого распределения записывается в следующем виде:

$$F_{2\text{mod}}(x) = F_1(x, \mu_1, \sigma_1, a_1, b_1) \cdot \alpha + F_2(x, \mu_2, \sigma_2, a_2, b_2) \cdot (1 - \alpha),$$

$$f_{2\text{mod}}(x) = \frac{dF_{2\text{mod}}(x)}{dx} = \frac{dF_1(x, \mu_1, \sigma_1, a_1, b_1)}{dx} \cdot \alpha + \frac{dF_2(x, \mu_2, \sigma_2, a_2, b_2)}{dx} \cdot (1 - \alpha), \quad (1.27)$$

где

μ_1 – математическое ожидание первой составляющей;

σ_1 – СКО первой составляющей;

a_1, b_1 – границы области рассеяния первой составляющей;

μ_2 – математическое ожидание второй составляющей;

σ_2 – СКО второй составляющей;

a_2, b_2 – границы области рассеяния второй составляющей;

α – доля первой составляющей в общем распределении,

или, используя обозначения Раздела 1.1.1.3:

$$F_{2\text{mod}}(x) = F_{LAD}(x, \mu_1, \sigma_1, a_1, b_1) \cdot \alpha + F_{LAD}(x, \mu_2, \sigma_2, a_2, b_2) \cdot (1 - \alpha), \quad (1.28)$$

$$f_{2\text{mod}}(x) = f_{LAD}(x, \mu_1, \sigma_1, a_1, b_1) \cdot \alpha + f_{LAD}(x, \mu_2, \sigma_2, a_2, b_2) \cdot (1 - \alpha). \quad (1.29)$$

Основные характеристики ДРОСП представлены в таблице 1.5.

Таблица 1.5. Характеристики ДРОСП

| Название характеристики | Формула или числовое значение характеристики |
|-------------------------|--|
| Обозначение | $LAD2(x; \mu_1, \sigma_1, a_1, b_1, \mu_2, \sigma_2, a_2, b_2, \alpha)$ |
| Параметры | $\mu_1, \sigma_1, a_1, b_1, \mu_2, \sigma_2, a_2, b_2, \alpha$ |
| Плотность распределения | $f_{LAD2}(x) = f_{LAD}(x, \mu_1, \sigma_1, a_1, b_1) \cdot \alpha + f_{LAD}(x, \mu_2, \sigma_2, a_2, b_2) \cdot (1 - \alpha)$ где $f_{LAD}(x)$ – плотность нормального распределения с ограниченной областью рассеяния $\mu_1, \sigma_1, a_1, b_1$ – параметры первой составляющей распределения, $\mu_2, \sigma_2, a_2, b_2$ – параметры второй составляющей распределения, α – доля первой составляющей в общем распределении |
| Функция распределения | $F_{2\text{mod}}(x) = F_{LAD}(x, \mu_1, \sigma_1, a_1, b_1) \cdot \alpha + F_{LAD}(x, \mu_2, \sigma_2, a_2, b_2) \cdot (1 - \alpha)$ |

Отметим, что такие основные параметры ДРОСП, как среднее значение и дисперсия, не могут быть получены аналитически в связи с наличием ограниченных областей рассеяния. Распределение (1.28) относится к классу

9-параметрических распределений. Проанализируем свойства функций (1.28) и (1.29), представленных на рисунках 1.22–1.27:

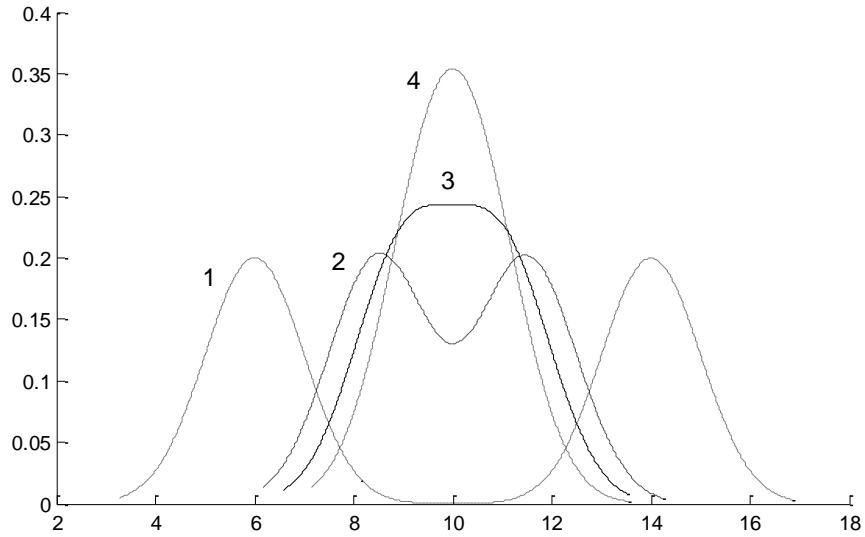


Рисунок 1.22. ДРОСП: график функции $f_{2\text{mod}}(x; \mu_1, \sigma_1, a_1, b_1, \mu_2, \sigma_2, a_2, b_2, \alpha)$

- 1 – $\mu_1 = 6, \mu_2 = 14, \sigma_1 = \sigma_2 = 1, a_1 = a_2 = 2, b_1 = b_2 = 18, \alpha = 0.5, x \in [a_1, b_2]$;
 2 – $\mu_1 = 8.5, \mu_2 = 11.5, \sigma_1 = \sigma_2 = 1, a_1 = a_2 = 3, b_1 = b_2 = 17, \alpha = 0.5, x \in [a_1, b_2]$;
 3 – $\mu_1 = 9, \mu_2 = 11, \sigma_1 = \sigma_2 = 1, a_1 = a_2 = 4, b_1 = b_2 = 16, \alpha = 0.5, x \in [a_1, b_2]$;
 4 – $\mu_1 = 9.5, \mu_2 = 10.5, \sigma_1 = \sigma_2 = 1, a_1 = a_2 = 5, b_1 = b_2 = 15, \alpha = 0.5, x \in [a_1, b_2]$

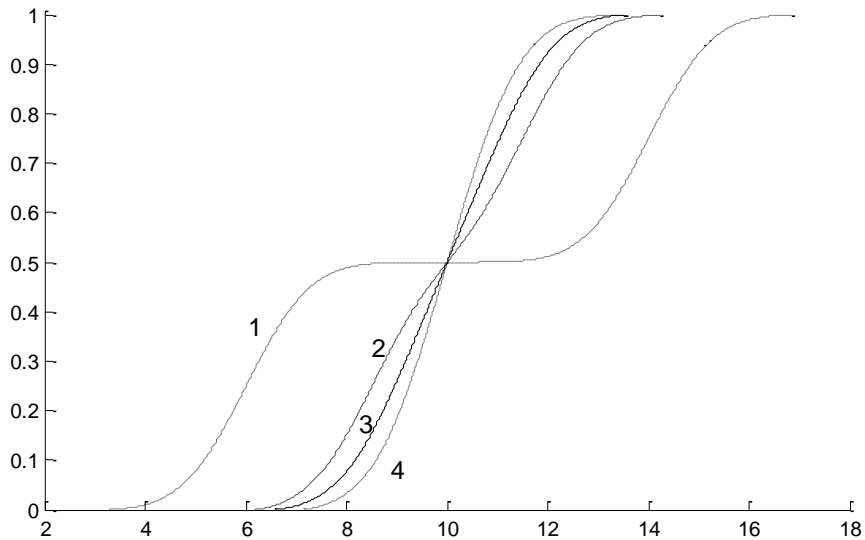


Рисунок 1.23. ДРОСП: график функции $F_{2\text{mod}}(x; \mu_1, \sigma_1, a_1, b_1, \mu_2, \sigma_2, a_2, b_2, \alpha)$

- 1 – $\mu_1 = 6, \mu_2 = 14, \sigma_1 = \sigma_2 = 1, a_1 = a_2 = 2, b_1 = b_2 = 18, \alpha = 0.5, x \in [a_1, b_2]$;
 2 – $\mu_1 = 8.5, \mu_2 = 11.5, \sigma_1 = \sigma_2 = 1, a_1 = a_2 = 3, b_1 = b_2 = 17, \alpha = 0.5, x \in [a_1, b_2]$;
 3 – $\mu_1 = 9, \mu_2 = 11, \sigma_1 = \sigma_2 = 1, a_1 = a_2 = 4, b_1 = b_2 = 16, \alpha = 0.5, x \in [a_1, b_2]$;
 4 – $\mu_1 = 9.5, \mu_2 = 10.5, \sigma_1 = \sigma_2 = 1, a_1 = a_2 = 5, b_1 = b_2 = 15, \alpha = 0.5, x \in [a_1, b_2]$

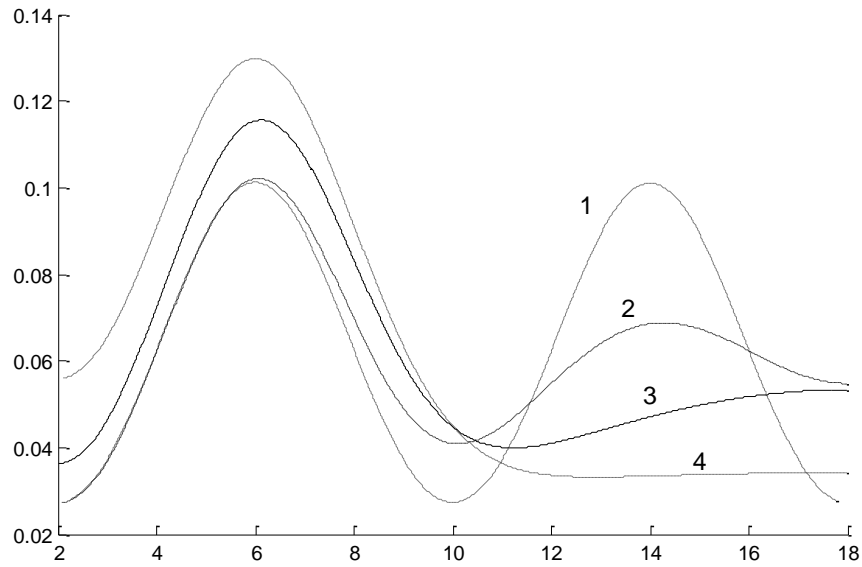


Рисунок 1.24. ДРОСП: график функции $f_{2\text{mod}}(x; \mu_1, \sigma_1, a_1, b_1, \mu_2, \sigma_2, a_2, b_2, \alpha)$

- 1 – $\mu_1 = 6, \mu_2 = 14, \sigma_1 = 2, \sigma_2 = 2, a_1 = a_2 = 2, b_1 = b_2 = 18, \alpha = 0.5, x \in [a_1, b_2]$;
- 2 – $\mu_1 = 6, \mu_2 = 14, \sigma_1 = 2, \sigma_2 = 3, a_1 = a_2 = 2, b_1 = b_2 = 18, \alpha = 0.5, x \in [a_1, b_2]$;
- 3 – $\mu_1 = 6, \mu_2 = 14, \sigma_1 = 2, \sigma_2 = 6, a_1 = a_2 = 2, b_1 = b_2 = 18, \alpha = 0.5, x \in [a_1, b_2]$;
- 4 – $\mu_1 = 6, \mu_2 = 14, \sigma_1 = 2, \sigma_2 = 12, a_1 = a_2 = 2, b_1 = b_2 = 18, \alpha = 0.5, x \in [a_1, b_2]$

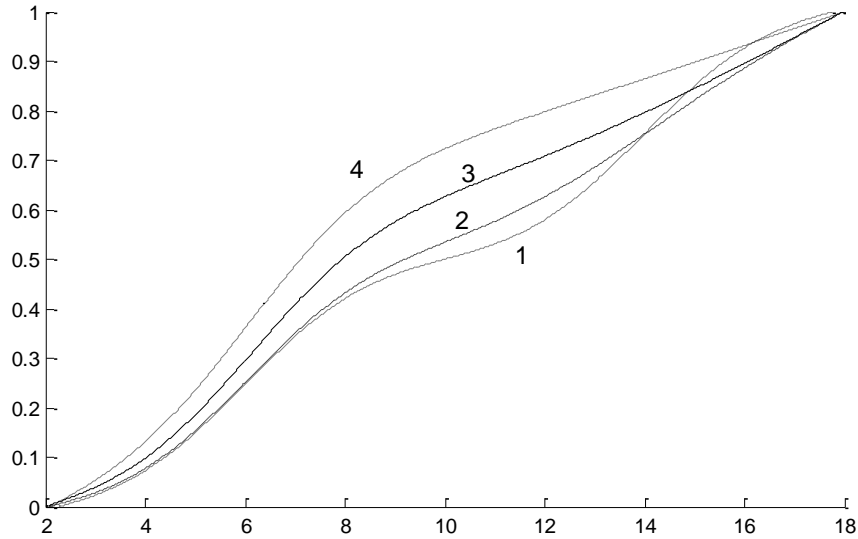


Рисунок 1.25. ДРОСП: график функции $F_{2\text{mod}}(x; \mu_1, \sigma_1, a_1, b_1, \mu_2, \sigma_2, a_2, b_2, \alpha)$

- 1 – $\mu_1 = 6, \mu_2 = 14, \sigma_1 = 2, \sigma_2 = 2, a_1 = a_2 = 2, b_1 = b_2 = 18, \alpha = 0.5, x \in [a_1, b_2]$;
- 2 – $\mu_1 = 6, \mu_2 = 14, \sigma_1 = 2, \sigma_2 = 3, a_1 = a_2 = 2, b_1 = b_2 = 18, \alpha = 0.5, x \in [a_1, b_2]$;
- 3 – $\mu_1 = 6, \mu_2 = 14, \sigma_1 = 2, \sigma_2 = 6, a_1 = a_2 = 2, b_1 = b_2 = 18, \alpha = 0.5, x \in [a_1, b_2]$;
- 4 – $\mu_1 = 6, \mu_2 = 14, \sigma_1 = 2, \sigma_2 = 12, a_1 = a_2 = 2, b_1 = b_2 = 18, \alpha = 0.5, x \in [a_1, b_2]$

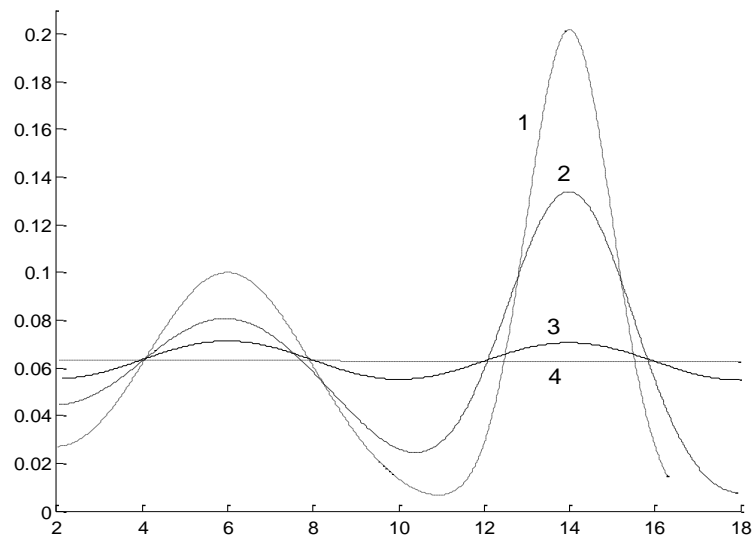


Рисунок 1.26. ДРОСП: график функции $f_{2\text{mod}}(x; \mu_1, \sigma_1, a_1, b_1, \mu_2, \sigma_2, a_2, b_2, \alpha)$

- 1 – $\mu_1 = 6, \mu_2 = 14, \sigma_1 = 2, \sigma_2 = 1, a_1 = a_2 = 2, b_1 = b_2 = 18, \alpha = 0.5, x \in [a_1, b_2]$;
 2 – $\mu_1 = 6, \mu_2 = 14, \sigma_1 = 2.5, \sigma_2 = 1.5, a_1 = a_2 = 2, b_1 = b_2 = 18, \alpha = 0.5, x \in [a_1, b_2]$;
 3 – $\mu_1 = 6, \mu_2 = 14, \sigma_1 = 3, \sigma_2 = 3, a_1 = a_2 = 2, b_1 = b_2 = 18, \alpha = 0.5, x \in [a_1, b_2]$;
 4 – $\mu_1 = 6, \mu_2 = 14, \sigma_1 = 6, \sigma_2 = 6, a_1 = a_2 = 2, b_1 = b_2 = 18, \alpha = 0.5, x \in [a_1, b_2]$

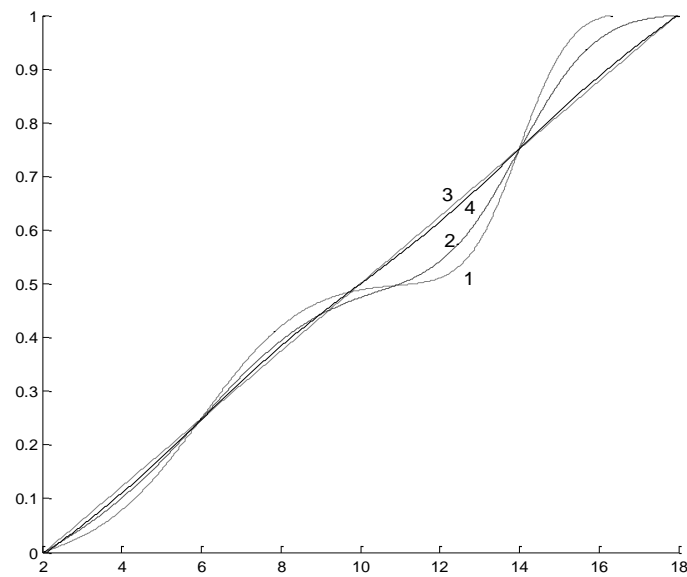


Рисунок 1.27. ДРОСП: график функции $F_{2\text{mod}}(x; \mu_1, \sigma_1, a_1, b_1, \mu_2, \sigma_2, a_2, b_2, \alpha)$

- 1 – $\mu_1 = 6, \mu_2 = 14, \sigma_1 = 2, \sigma_2 = 1, a_1 = a_2 = 2, b_1 = b_2 = 18, \alpha = 0.5, x \in [a_1, b_2]$;
 2 – $\mu_1 = 6, \mu_2 = 14, \sigma_1 = 2.5, \sigma_2 = 1.5, a_1 = a_2 = 2, b_1 = b_2 = 18, \alpha = 0.5, x \in [a_1, b_2]$;
 3 – $\mu_1 = 6, \mu_2 = 14, \sigma_1 = 3, \sigma_2 = 3, a_1 = a_2 = 2, b_1 = b_2 = 18, \alpha = 0.5, x \in [a_1, b_2]$;
 4 – $\mu_1 = 6, \mu_2 = 14, \sigma_1 = 6, \sigma_2 = 6, a_1 = a_2 = 2, b_1 = b_2 = 18, \alpha = 0.5, x \in [a_1, b_2]$

Из рисунков 1.22–1.27 можно сделать следующие выводы:

- При $\frac{|\mu_1 - \mu_2|}{\sqrt{2(\sigma_1^2 + \sigma_2^2)}} > 1$ у функции ПР можно выделить две моды.
- При $\frac{|\mu_1 - \mu_2|}{\sqrt{2(\sigma_1^2 + \sigma_2^2)}} > 4$ части распределения практически не влияют

друг на друга и функция плотности суммарного распределения имеет две ярко выраженные моды.

- При $\frac{|\mu_1 - \mu_2|}{\sqrt{2(\sigma_1^2 + \sigma_2^2)}} < 1$ становится труднее определить абсциссы мод

составляющих распределения, и распределение становится похожим на одномодальное нормальное распределение;

- При $\frac{|b_2 - a_1|}{\sqrt{2(\sigma_1^2 + \sigma_2^2)}} > 3$ ФР приближается по форме к ФР равномерного

распределения.

1.2 Постановка задачи оценки функций распределения случайных последовательностей с ограниченной областью рассеяния и анализ методов ее решения

Восстановление ФР по выборке случайных данных, полученных в результате проведения тех или иных экспериментов, является основной задачей математической статистики [11,17,25,31,35,39]. Задача имеет важное практическое значение, например, при решении задач прочностной надежности элементов [41,115] и объектов нефтегазового оборудования [45,74,86].

Данная задача имеет следующую постановку: по экспериментальной выборке значений случайной величины $X_i, i = \overline{1, N_s}$ из генеральной совокупности найти функцию распределения $F(y) = \Pr\{X \leq y\}$, связанную с плотностью распределения $f(y)$ интегральным соотношением

$$F(y) = \int_{-\infty}^y f(\xi) d\xi. \quad (1.30)$$

Известны два основных подхода к решению этой задачи: параметрический и непараметрический.

1.2.1 Параметрический подход

Параметрический подход предусматривает выбор на основе той априорной информации вида ФР случайной величины $F(y)$, зависящей от некоторого набора параметров, и получении оценок их значений по

имеющейся выборке данных, обеспечивающих максимальную близость теоретической ФР $F(y)$ и эмпирической ФР $F_{Ns}(y)$

$$F_{Ns}(y) = \frac{1}{Ns} \sum_{i=1}^{Ns} \Theta(y - x_i), \quad (1.31)$$

где функция Хэвисайда

$$\Theta(y - x_i) = \begin{cases} 1, & \text{при } y - x_i \geq 0, \\ 0, & \text{при } y - x_i < 0, \end{cases}$$

в соответствие с выбранной мерой близости, зависящей, вообще говоря, от вида ФР [54].

Существование решения обсуждаемой задачи обеспечивает центральная теорема математической статистики (теорема Гливенко) [8], согласно которой с ростом объема выборки Ns функция $F_{Ns}(y)$ с вероятностью, равной единице, равномерно приближается к $F(y)$:

$$\Pr \left\{ \lim_{Ns \rightarrow \infty} \sup |F_{Ns}(y) - F(y)| = 0 \right\} = 1.$$

В общем случае неизвестная истинная ФР имеет вид $F(y, \bar{a})$, где \bar{a} – вектор параметров. Традиционно, вектор параметров \bar{a} находят как решение следующей задачи:

$$\bar{a} = \arg \min_a \left(\Phi(F(y, a) - F_{Ns}(y)) \right), \quad (1.32)$$

где Φ – тот или иной функционал, вид которого зависит от выбранного критерия близости функций $F(y, \bar{a})$ и $F_{Ns}(y)$. Заменяв в (1.32) $F(y, \bar{a})$, $F_{Ns}(y)$ на $f(y, \bar{a})$, $f_{Ns}(y)$, где значения функции $f_{Ns}(y)$ есть соответствующие частоты гистограммы случайной последовательности, приходим к следующей постановке задачи оценивания параметров:

$$[\bar{x}_0, \bar{\sigma}, \bar{x}_{\min}, \bar{x}_{\max}] = \arg \min_{x_0, \sigma, x_{\min}, x_{\max}} \left(\Phi(F_{LAD}(y; x_0, \sigma, x_{\min}, x_{\max}) - F_{Ns}(y)) \right), \quad (1.33)$$

$$[\bar{x}_0, \bar{\sigma}, \bar{x}_{\min}, \bar{x}_{\max}] = \arg \min_{x_0, \sigma, x_{\min}, x_{\max}} \left(\Phi(f_{LAD}(y; x_0, \sigma, x_{\min}, x_{\max}) - f_{Ns}(y)) \right). \quad (1.34)$$

Применительно к рассматриваемой в данном исследовании задачи оценивания параметров случайных распределений с ограниченной областью рассеяния задачи (1.32), принимают следующий вид:

$$[\bar{x}_0, \bar{\sigma}, \bar{x}_{\min}, \bar{x}_{\max}] = \arg \min \left(\Phi(F_{LAD}(y; x_0, \sigma, x_{\min}, x_{\max}) - F_{Ns}(y)) \right) = \arg \min_{x_0, \sigma, x_{\min}, x_{\max}} \left(\Phi \left(\int_{\bar{x}_{\min}}^y f_{LAD}(\xi, x_0, \sigma, x_{\min}, x_{\max}) d\xi - F_{Ns}(y) \right) \right), \quad (1.35)$$

$$\begin{aligned}
[\bar{x}_0, \bar{\sigma}, \bar{x}_{\min}, \bar{x}_{\max}] = \arg \min_{x_0, \mu, x_{\min}, x_{\max}} & \left(\Phi(f_{LAD}(y; x_0, \sigma, x_{\min}, x_{\max}) - f_{Ns}(y)) \right) = \\
\arg \min_{x_0, \sigma, x_{\min}, x_{\max}} & \left(\Phi \left(\exp \left[-(y - x_0)^2 / 2\sigma^2 \right] + \sum_{g=0}^{\infty} \exp \left[-(y - x_{2g+1}^{\pm})^2 / 2\sigma^2 \right] + \right. \right. \\
& \left. \left. \sum_{g=1}^{\infty} \exp \left[-(y - x_{2g}^{\pm})^2 / 2\sigma^2 \right] - f_{Ns}(y) \right) \right), \quad (1.36)
\end{aligned}$$

где $x_{2g}^{\pm} = \pm 4g(x_{\max} - x_{\min}) + x_0$, $x_{2g+1}^{\pm} = \pm(4g+2)(x_{\max} - x_{\min}) - x_0$, $g = 0, 1, \dots$

Из (1.35), (1.36) видно, что задача оценки параметров изучаемого распределения оказываются, как в первой, так и во второй постановках, оказываются нелинейными. В этой связи для их решения можно, например, использовать методы оптимизации или численные методы решения систем нелинейных уравнений, например, метод Ньютона и его модификации, методы нелинейного программирования и т.д. Однако точность оценки параметров с помощью данных методов оказывается существенно зависящей от выбора начальных значений оцениваемых параметров. Это иллюстрируют результаты статистического моделирования, в котором генерировались $numExp = 100$ выборок случайных чисел объемом $Ns = 1000$ случайных чисел, распределенных в соответствии с законом $F_{LAD}(x, 60, 30, 0, 100)$ (использовался метод обратной функции [1,8]). Далее для каждой из случайных реализаций находились оценки параметров LAD -распределения (ансамбли оценок значений параметров распределения $F_{LAD}(x, 60, 30, 0, 100)$) с помощью метода Нелдера-Мида. Усредненные по ансамблям значения погрешности оценок и соответствующие начальные значения параметров $\bar{x}_0, \bar{\sigma}, \bar{x}_{\min}, \bar{x}_{\max}$ представлены в таблице 1.6.

Таблица 1.6 Погрешность оценок параметров распределения $F_{LAD}(x, 60, 30, 0, 100)$ (полужирным выделены неудовлетворительные результаты)

| Начальные значения | | | | Погрешности оценок параметров, % | | | |
|--------------------|----------------|------------------|------------------|--|--|-------------------------------------|---------------------------------------|
| \bar{x}_0 | $\bar{\sigma}$ | \bar{x}_{\min} | \bar{x}_{\max} | Δx_0 | $\Delta \sigma$ | Δx_{\min} | Δx_{\max} |
| 10 | 10 | 0 | 100 | $2,45 \cdot 10^{10}$ | $2,52 \cdot 10^{10}$ | $1,25 \cdot 10^7$ | $6,3 \cdot 10^{10}$ |
| 20 | 10 | 0 | 100 | $2,61 \cdot 10^0$ | $1,69 \cdot 10^0$ | $4,46 \cdot 10^0$ | $8,63 \cdot 10^4$ |
| 10 | 20 | 0 | 100 | $1,49 \cdot 10^4$ | $1,95 \cdot 10^0$ | $0,17 \cdot 10^0$ | $61,60 \cdot 10^0$ |
| 40 | 40 | 0 | 100 | $0,92 \cdot 10^0$ | $0,64 \cdot 10^0$ | $0,87 \cdot 10^0$ | $4,85 \cdot 10^0$ |
| 60 | 10 | 0 | 100 | $1,61 \cdot 10^0$ | $1,48 \cdot 10^0$ | $0,31 \cdot 10^0$ | $7,00 \cdot 10^3$ |
| 80 | 50 | 0 | 100 | $5,62 \cdot 10^0$ | $2,89 \cdot 10^0$ | $0,20 \cdot 10^{-1}$ | $4,82 \cdot 10^0$ |
| 90 | 90 | 0 | 100 | $58,09 \cdot 10^0$ | $5,14 \cdot 10^0$ | $0,20 \cdot 10^{-1}$ | $3,85 \cdot 10^0$ |

Из таблицы 1.6 видно, что точность оценок параметров распределения оказывается существенно зависящей от выбора их начального приближения, которое на практике в большинстве случаев, априори, далеко не очевидно. К тому же при использовании численных методов не всегда удается

реализовать ограничение области поиска параметров, что приводит к нахождению абсолютно непригодных решений. В этой связи требуется привлечение методов решения задач (1.35), (1.36) свободных от указанных недостатков.

Отметим, что при наличии априорной информации, позволяющей выбрать вид ПР $f(x, \Theta)$, $\Theta = \{\Theta_1, \Theta_2, \dots, \Theta_N\}$ – вектор параметров, СП x_i , $i = \overline{1, N}$, для оценки ее параметров также можно использовать метод максимального правдоподобия [1,8], предложенный Фишером. Данный метод состоит в том, что в качестве наиболее правдоподобного значения параметров выбирают $\bar{\Theta}$, координаты которого максимизируют вероятность получения при N испытаниях выборку $x = \{x_1, x_2, \dots, x_N\}$ и, соответственно, функцию правдоподобия $L(f(x, \Theta), x_i)$. Напомним, что в случае непрерывной ПР $f(x, \Theta)$ функция правдоподобия $L(f(x, \Theta), x_i)$ описывается совместной ПР:

$$L(f(x, \Theta), x_i) = f(x_1, \Theta) f(x_2, \Theta) \cdot \dots \cdot f(x_N, \Theta), \quad (1.37)$$

в случае дискретного распределения:

$$L(f(X, \Theta), x_i) = P_{\Theta}(X_1 = x_1) P_{\Theta}(X_2 = x_2) \cdot \dots \cdot P_{\Theta}(X_N = x_N). \quad (1.38)$$

При наблюдаемых значениях $x = \{x_1, x_2, \dots, x_N\}$ функция $L(f(x, \Theta), x_i)$ является функцией вектора параметров Θ , поэтому оценка максимального правдоподобия $\bar{\Theta}$, есть решение следующей задачи:

$$\bar{\Theta} = \arg \max_{\Theta} L(f(x_i, \Theta), x_i). \quad (1.39)$$

Из (1.39) видно, что оценки максимального правдоподобия удовлетворяют уравнению

$$\frac{\partial L}{\partial \Theta} = 0 \quad (1.40)$$

или

$$\frac{\partial \ln L}{\partial \Theta} = 0. \quad (1.41)$$

Подставив выражение для плотности распределения $f_{LAD}(x, x_0, \sigma, x_{\min}, x_{\max})$ в (1.37) и вычислив его натуральный логарифм, получаем

$$\begin{aligned}
\ln L = & \exp\left[-(x_1 - x_0)^2 / 2\sigma^2\right] + \sum_{g=0}^{\infty} \exp\left[-(x_1 - x_{2g+1}^{\pm})^2 / 2\sigma^2\right] + \\
& \sum_{g=1}^{\infty} \exp\left[-(x_1 - x_{2g}^{\pm})^2 / 2\sigma^2\right] + \\
& \dots \\
& \exp\left[-(x_N - x_0)^2 / 2\sigma^2\right] + \sum_{g=0}^{\infty} \exp\left[-(x_N - x_{2g+1}^{\pm})^2 / 2\sigma^2\right] + \\
& \sum_{g=1}^{\infty} \exp\left[-(x_N - x_{2g}^{\pm})^2 / 2\sigma^2\right].
\end{aligned} \tag{1.42}$$

Продифференцировав (1.42) по параметрам и принимая во внимание, что

$$\begin{aligned}
\frac{\partial x_{2g}^{\pm}}{\partial x_0} = 1, \quad \frac{\partial x_{2g}^{\pm}}{\partial x_{\max}} = \pm 4g, \quad \frac{\partial x_{2g}^{\pm}}{\partial x_{\min}} = \mp 4g, \\
\frac{\partial x_{2g+1}^{\pm}}{\partial x_0} = -1, \quad \frac{\partial x_{2g+1}^{\pm}}{\partial x_{\max}} = \pm(4g+2), \quad \frac{\partial x_{2g+1}^{\pm}}{\partial x_{\min}} = \mp(4g+2),
\end{aligned}$$

получаем следующую систему уравнений:

$$\begin{aligned}
\frac{\partial L}{\partial x_0} = & \left[(x_1 - x_0) / \sigma^2\right] \exp\left[-(x_1 - x_0)^2 / 2\sigma^2\right] + \\
& - \sum_{g=0}^{\infty} \left[(x_1 - x_{2g+1}^{\pm}) / \sigma^2\right] \exp\left[-(x_1 - x_{2g+1}^{\pm})^2 / 2\sigma^2\right] + \\
& + \sum_{g=1}^{\infty} \left[(x_1 - x_{2g}^{\pm}) / \sigma^2\right] \exp\left[-(x_1 - x_{2g}^{\pm})^2 / 2\sigma^2\right] + \dots = 0, \\
\frac{\partial L}{\partial \sigma} = & \left[(x_1 - x_0)^2 / \sigma^3\right] \exp\left[-(x_1 - x_0)^2 / 2\sigma^2\right] + \\
& \sum_{g=0}^{\infty} \left[(x_1 - x_{2g+1}^{\pm})^2 / \sigma^3\right] \exp\left[-(x_1 - x_{2g+1}^{\pm})^2 / 2\sigma^2\right] + \\
& \sum_{g=1}^{\infty} \left[(x_1 - x_{2g}^{\pm}) / \sigma^3\right] \exp\left[-(x_1 - x_{2g}^{\pm})^2 / 2\sigma^2\right] + \dots = 0,
\end{aligned} \tag{1.43}$$

(Начало)

$$\begin{aligned}
\frac{\partial L}{\partial x_{\min}} &= \sum_{g=0}^{\infty} \left[\mp (4g+2)(x_1 - x_{2g+1}^{\pm}) / \sigma^2 \right] \exp \left[-(x_1 - x_{2g+1}^{\pm})^2 / 2\sigma^2 \right] + \\
&\sum_{g=1}^{\infty} \left[\mp 4g(x_1 - x_{2g}^{\pm}) / \sigma^2 \right] \exp \left[-(x_1 - x_{2g}^{\pm})^2 / 2\sigma^2 \right] + \dots = 0, \\
\frac{\partial L}{\partial x_{\max}} &= \sum_{g=0}^{\infty} \left[\pm (4g+2)(x_1 - x_{2g+1}^{\pm}) / \sigma^2 \right] \exp \left[-(x_1 - x_{2g+1}^{\pm})^2 / 2\sigma^2 \right] + \\
&\sum_{g=1}^{\infty} \left[\pm 4g(x_1 - x_{2g}^{\pm}) / \sigma^2 \right] \exp \left[-(x_1 - x_{2g}^{\pm})^2 / 2\sigma^2 \right] + \dots = 0.
\end{aligned} \tag{1.43}$$

Продолжение

Сравнивая два последних уравнения в (1.43), видим, что их сумма оказывается равной нулю, т.е. данные уравнения оказываются тождественными друг другу. Следовательно, для рассматриваемого распределения метод максимального правдоподобия приводит к недоопределенной системе уравнений. Такие системы могут, вообще говоря, может иметь бесконечное число решений, для выбора из которых единственно верного требуется, использование дополнительной информации, которая изначально в рассматриваемой задаче отсутствует. В этой связи, можно поставить под сомнение возможность использования метода максимального правдоподобия для оценки параметров СПООР.

1.2.2 Непараметрический подход

В основе непараметрической статистики лежит подход, позволяющий получать адаптивные оценки эмпирических распределений в виде некоторых функционалов, независимых от вида неизвестного априорного распределения [72,75,77,79]. Для восстановления неизвестной ФР в непараметрической статистике разработан целый ряд методов и алгоритмов [72]: метод гистограмм, «гребенка», метод ближайших соседей, метод разложения по базисным функциям, аппроксимация Розенблатта-Парзена и ряд других. Работоспособность методов непараметрической статистики и целесообразность их применения при анализе экспериментальных данных подтверждается результатами, полученными различными исследователями [72].

Напомним, следуя [74], что метод Розенблатта-Парзена восстановления (аппроксимации) ФР экспериментальной выборки основан на предположении о том, что ФР оценивается локально в каждой точке x_i с помощью элементов обучающей выборки из некоторой окрестности x_i . При этом общая ФР $F(y)$ есть сумма локальных функций

$$F(y) = \frac{1}{N_s} \sum_{i=1}^{N_s} K\left(\frac{y - x_i}{h}\right), \tag{1.44}$$

где $K(t)$ – ядерная функция, удовлетворяющая следующим условиям:

а) $K(t)$ – монотонно неубывающая функция, область значений которой принадлежит интервалу $[0,1]$;

б) $K(t) = 1 - K(t)$ – функция, симметричная относительно 0;

в) $h \rightarrow 0$ при $Ns \rightarrow \infty$;

h – параметр «размытости», определяющий гладкость получаемой оценки.

В качестве ядерных функций $k(y)$, традиционно, используют функции, представленные в таблице 1.7.

Таблица 1.7. Функции, используемые в качестве ядерных функций

| № | Ядро | Формула |
|---|---------------|--|
| 1 | Нормальное | $k(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$ |
| 2 | Лапласа | $k(t) = \frac{1}{2} e^{- t }$ |
| 3 | Фишера | $k(t) = \frac{1}{2\pi} \left(\frac{\sin\left(\frac{t}{2}\right)}{\frac{t}{2}} \right), \left \frac{t}{2} \right \leq \pi$ |
| 4 | Коши | $k(t) = \frac{1}{\pi} \left(\frac{1}{1+t^2} \right)$ |
| 5 | Логистическое | $k(t) = \frac{e^{-t}}{(1+e^{-t})^2}$ |
| 6 | Епанечникова | $k(t) = \frac{3 \cdot \left(1 - \frac{t^2}{5}\right)}{4\sqrt{5}}, t \leq \sqrt{5}$ |
| 7 | Равномерное | $k(t) = \frac{1}{2}, t \leq 1$ |
| 8 | Треугольное | $k(t) = 1 - t , t \leq 1$ |
| 9 | Квадратичное | $k(t) = \frac{3 \cdot (1-t^2)}{4}, t \leq 1$ |

Соответственно, плотности вероятности $f(y)$ вычисляется по формуле

$$f(y) = \frac{1}{Ns \cdot h} \sum_{i=1}^{Ns} k\left(\frac{y - x_i}{h}\right), \quad (1.45)$$

где

$$k(y) = \frac{d}{dy} K(y).$$

Данные оценки предложены Розенблаттом [121] и исследованы Парзеном [117].

В рассматриваемом методе качество аппроксимации зависит от вида ядра $k(t)$ и от значения параметра размытости h [72] (рисунок 1.28).

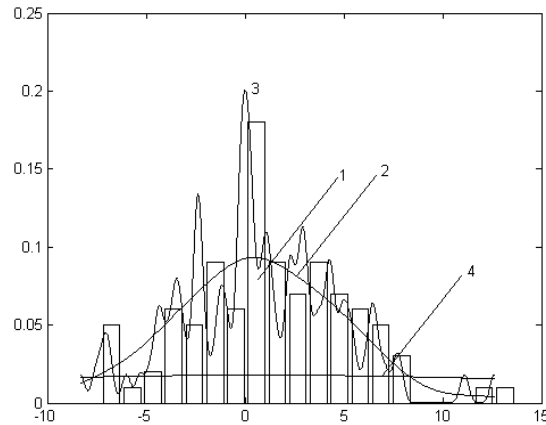


Рисунок 1.28. ПР случайной последовательности $x_i, i = \overline{1, 100}$, сгенерированной в соответствии с НР $N(1, 4)$: 1 – гистограмма случайной последовательности; 2 – нормальное ядро, $h = h_1^*$, где h_1^* пока не определено; 3 – нормальное ядро, $h < h_1^*$, где h_1^* пока не определено; 4 – нормальное ядро, $h > h_1^* > h_1^*$, где h_1^* пока не определено

Оптимальные значения ядерной функции и параметра находятся из условия достижения функционалом

$$J = \int \ln k(t) \cdot f(t) dt \quad (1.46)$$

максимального значения, которое, как очевидно, выполняется при $k(t) = f(t)$.

При анализе экспериментальных данных задача поиска оптимального значения параметра «размытости» в [74] рекомендовано находить для каждой из ядерных функций, представленных в таблице 1.7, оптимальное значения $h_m^*, m = \overline{1, 9}$, как решение задачи

$$h_m^* = \arg \max \left\{ \frac{1}{N_s} \sum_{i=1}^{N_s} \ln \left[\frac{1}{(N_s - 1) \cdot h_m} \sum_{j \neq i}^{N_s-1} k_m \left(\frac{x_i - x_j}{h_m} \right) \right] \right\}, \quad (1.47)$$

и далее выбирать из них ту ядерную функцию $k_m(y)$, для которой значение функции

$$\varphi(h_m) = \frac{1}{N_s} \sum_{i=1}^{N_s} \ln \left[\frac{1}{(N_s - 1) \cdot h_m} \sum_{j \neq i}^{N_s-1} k_m \left(\frac{x_i - x_j}{h_m} \right) \right] \quad (1.48)$$

в точке $h_m = h_m^*$ будет наибольшим.

Из (1.47) видно, что нахождение оптимального значения параметра размытости h для каждой из базисных функций сводится к решению сложного нелинейного уравнения

$$\frac{\partial \varphi(h)}{\partial h} = \frac{\partial}{\partial h} \sum_{i=1}^{N_s} \ln \left[\frac{1}{(N_s - 1) \cdot h} \sum_{j \neq i}^{N_s-1} k_m \left(\frac{x_i - x_j}{h} \right) \right] = 0, \quad (1.49)$$

найти которое оказывается возможным только численно.

Таким образом, в общем случае решение описанных выше задач (1.32), (1.44), (1.47) приводит к решению системы нелинейных уравнений. Как известно, для системы нелинейных уравнений не существует универсального детерминированного метода решения, поэтому применяют численные методы, например, итерационный метод Ньютона, метод градиентного спуска, симплекс-метод [2,5,6,13,14,27,42,44,71,76,78,102,112,116]. Однако, известно [30,84,108,109,110], что для итерационных методов сходимость к истинному решению очень сильно зависит от начального приближения. В этой связи представляются перспективными эвристические методы случайного поиска, результативность которых, как утверждается, не зависит от начального приближения и позволяет найти оптимальное решение при любых начальных условиях. Одним из таких методов являются ГА [22,85].

1.3 Постановка задач исследования

Обоснована целесообразность использования в задаче аппроксимации ПР и ФР СПООР математической модели, предложенной А. Эйнштейном и М. Смолуховским. В соответствие с данной моделью, построенной на основе анализа случайных броуновских блужданий в ограниченной области рассеяния, ПР и ФР представляются линейными комбинациями нормальных законов распределения с центрами в точках нахождения зеркальных отражений точки, в которой броуновская частица находилась в момент времени $t = 0$ (центра рассеяния) и ее образов, относительно границ области рассеяния.

Показано, что использование классических методов оценивания параметров распределений СПООР встречает определенные трудности:

1. при использовании метода максимального правдоподобия получается недоопределенная система нелинейных уравнений, которая, как правило, имеет бесконечно много решений, выбор из которых единственно верного требует привлечений дополнительной информации о свойствах изучаемой случайной величины;

2. при использовании оптимизационных методов возникает необходимость решения той или иной системы нелинейных уравнений, точность которых оказывается зависящей от выбранного начального приближения.

Один из возможных путей выхода из данной ситуации состоит в использовании эвристических методов случайного поиска, сходимость которых, как утверждается, не зависит от начального приближения и, следовательно, точное решение, потенциально, может быть найдено при любых начальных условиях, одним из которых являются ГА. Однако универсальной методики использования ГА для решения задач параметрической статистики, пригодной для различных ПР и ФР, не существует.

В этой связи представляется целесообразным проведение целенаправленных исследований с целью изучения особенностей использования ГА в задаче оценивания параметров ФР и ПР СПООР.

Цель исследования: разработать научно-обоснованное математическое и алгоритмическое обеспечение для использования ГА в задаче оценки параметров распределений случайных последовательностей с ограниченными областями рассеяния на основе подходов параметрической и непараметрической статистики.

Для достижения поставленной цели необходимо решить следующие задачи:

1. Разработать и обосновать математическое и алгоритмическое обеспечение для оценки параметров распределений СПООР, обеспечивающее:

- оценку параметров одномодальных распределений с ограниченной областью рассеяния на основе ГА;
- оценку параметров двумодальных случайных величин с ограниченной областью рассеяния на основе ГА и аппроксимации Розенблатта-Парзена.

2. Разработать программную библиотеку, обеспечивающую практическое использование математического и алгоритмического обеспечения для оценки параметров распределений СПООР.

3. Применить разработанное математическое и алгоритмическое обеспечение для оценки параметров распределений СПООР.

Equation Chapter (Next) Section 1

2 Разработка математического и алгоритмического обеспечения для оценивания параметров одномодальных распределений случайных последовательностей с ООР

В данной главе рассмотрены следующие вопросы:

- Постановка задачи оценивания параметров одномодальных ФР и ПР СПООР в терминологии ГА.
- Анализ особенностей одномодальных ФР и ПР СПООР с ограниченной областью рассеяния.
- Обоснование рекомендаций по выбору настроек ГА при оценивании параметров одномодальных ФР и ПР СПООР.
- Анализ точности оценок параметров одномодальных ФР и ПР СПООР с помощью ГА.
- Обоснование математического и алгоритмического обеспечения для оценивания параметров одномодальных ФР и ПР СПООР на основе ГА.

2.1 Постановка задачи оценивания параметров одномодальных распределений случайных последовательностей с помощью ГА

Модели одномодальных ПР (1.14) и ФР (1.16) СПООР, построенные в Главе 1 в соответствие с методом мнимых источников, являются 5-ти параметрическими. Однако, как показано в разделе 1.1.1.3, вместо бесконечной системы мнимых источников оказывается достаточным задать в (1.14), (1.16) $N = 5$, то есть использовать $2N = 10$ мнимых источников:

$$f_{LAD}(x; \mu, \sigma, x_{\min}, x_{\max}, 5) = A \left[\begin{aligned} &\varphi(x; \mu, \sigma, x_{\min}, x_{\max}) + \\ &+ \sum_{g=0}^5 \varphi_{2g+1}^{\pm}(x; \mu, \sigma, x_{\min}, x_{\max}) \\ &+ \sum_{g=1}^5 \varphi_{2g}^{\pm}(x; \mu, \sigma, x_{\min}, x_{\max}) \end{aligned} \right], \quad (2.1)$$

$$F_{LAD}(x; \mu, \sigma, x_{\min}, x_{\max}, 5) = \int_a^x f_{LAD}(\xi; \mu, \sigma, x_{\min}, x_{\max}, 5) d\xi, \quad (2.2)$$

Таким образом, одномодальные ПР и ФР СПООР, и, соответственно, решаемая оптимизационная задача оказываются 4-х параметрическими.

Рассмотрим обобщенную блок-схему работы классического ГА, представленную на рисунке 2.1 [16,70,93,94,95,120].

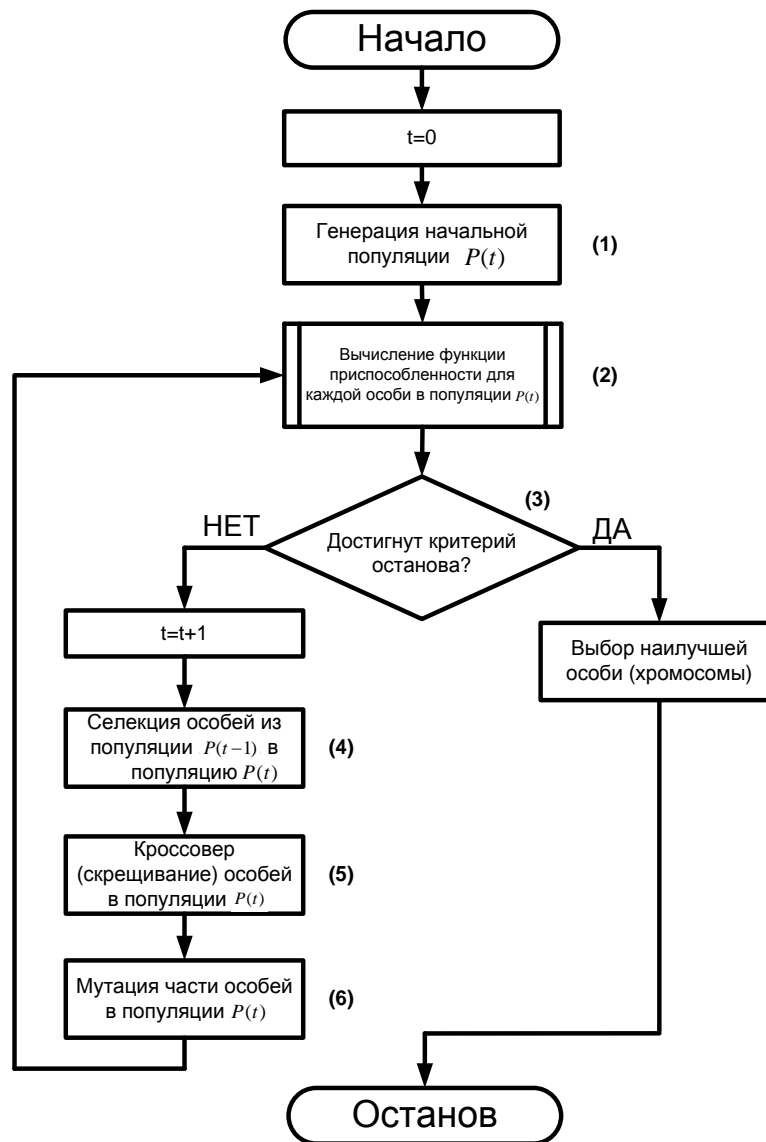


Рисунок 2.1. Блок схема классического ГА

Из рисунка 2.1 видно, что для применения ГА необходимо установить взаимно однозначное соответствие между основными структурными элементами ГА: видом элемента популяции (особи), оператор кроссовера (скрещивания), мутации, селекции (отбора), видом функции приспособленности (фитнес-функции) и параметрами изучаемого распределения $\mu, \sigma, x_{\min}, x_{\max}, N = 5$.

Напомним, что в ГА используется следующая иерархия понятий [96,101,122,124]: Популяция, состоящая из особей; **особь**, состоящая из одной или нескольких хромосом; **хромосомы**, содержащие закодированное текущее решение рассматриваемой задачи (в данном случае, значения параметров распределения). Хромосома состоит из последовательности генов, в дискретном случае имеющих значения 0 или 1.

Рассматриваемая задача относится к классу оптимизационных задач в непрерывной области поиска [91,99,100,127], поэтому каждая хромосома представляется 5-х мерным вектором, координаты которого (гены) имеют

вещественные значения, соответствующие параметрам ПР (2.1). Соответственно, отдельная особь в популяции состоит из одной хромосомы и пяти генов, один из которых остается неизменным (рисунок 2.2а).

Далее, следуя общей блок-схеме ГА, решение рассматриваемой задачи находилось выполнением следующей последовательности действий.

1. Формирование начальной популяции, состоящей из $popSize$ особей (блок 1 на рисунке 2.1). (Отметим, что число особей определяет производительность ГА.) С математической точки зрения популяция есть матрица вещественных чисел P , размерности $popSize \times 5$, в которой каждая строка-вектор – отдельная особь (рисунок 2.2б).

Начальная популяция в ГА, как правило, инициализируется случайным образом [89]: каждому гену присваивается случайное число, выбираемое из генеральной совокупности с известной функцией распределения. На практике обычно используются случайные величины с равномерным распределением в ограниченной области. Следовательно, в выбранном методе определение области ограничения значений генов также является важной задачей.

2. Вычисление для каждой особи в популяции значений **функции приспособленности** (ФП) особей (блок 2 на рисунке 2.1)

$$\Phi(P_i) = \Phi(\{\mu_i, \sigma_i, x_{\min i}, x_{\max i}, 5\}), \quad (2.3)$$

где P_i — i -ая особь в популяции P .

ФП (2.3) позволяет количественно оценить особь, сравнив ее с неким эталоном. Значения ФП используются для количественной оценки пригодности данного решения: если $\Phi(P_i) < \Phi(P_j)$, то особь P_i более приспособлена (лучше) особи P_j .

3. Проведение **селекции** (отбора) популяции [88] (Блок 4 на рисунке 2.1). **Селекция** – процесс, посредством которого более приспособленные особи получают большую возможность для воспроизводства потомков.

Основные виды селекции:

- *Селекция на основе рулетки.* При его реализации каждой особи в популяции соответствует зона на колесе рулетки, пропорционально соразмерная с величиной функции приспособленности. Тогда при повороте колеса рулетки каждая особь имеет некоторую вероятность выбора для селекции.

- *Элитная селекция.* В этом случае выбираются лучшие (элитные) особи на основе сравнения значений ФП.

- *Турнирная селекция.* Согласно размеру «турнира» случайно выбирается некоторое число особей, и лучшие особи в этой группе выбираются для воспроизведения.

• *Равномерная селекция.* Все особи имеют равные шансы для воспроизведения.

В результате выполнения селекции из популяции P размером $popSize$ особей формировалась временная «родительская» популяция P' в количестве $2 \cdot popSize$ особей, сгруппированных по парам. При этом родительскую популяцию образуют выделенные более приспособленные особи (рисунок 2.2в).

4. Применение к каждой паре родителей оператора **кроссовера** (блок 5 на рисунке 2.1). **Кроссовер (скрещивание)** – процесс, позволяющий на основе родительских особей получать потомков.

Основными видами кроссовера являются:

- **Одноточечный кроссовер.** При его использовании в двух случайных хромосомах случайным образом выбирается точка, в которой они должны быть «разрезаны». Далее меняют местами части хромосом, расположенные справа от точки кроссовера получают двух новых потомков (рисунок 2.2г).
- **Двухточечный кроссовер.** При его использовании случайным образом в каждой хромосоме выбираются две точки кроссовера и хромосомы обмениваются соответствующими генами.
- **Многоточечный кроссовер.** Выполняется аналогично выше описанному.

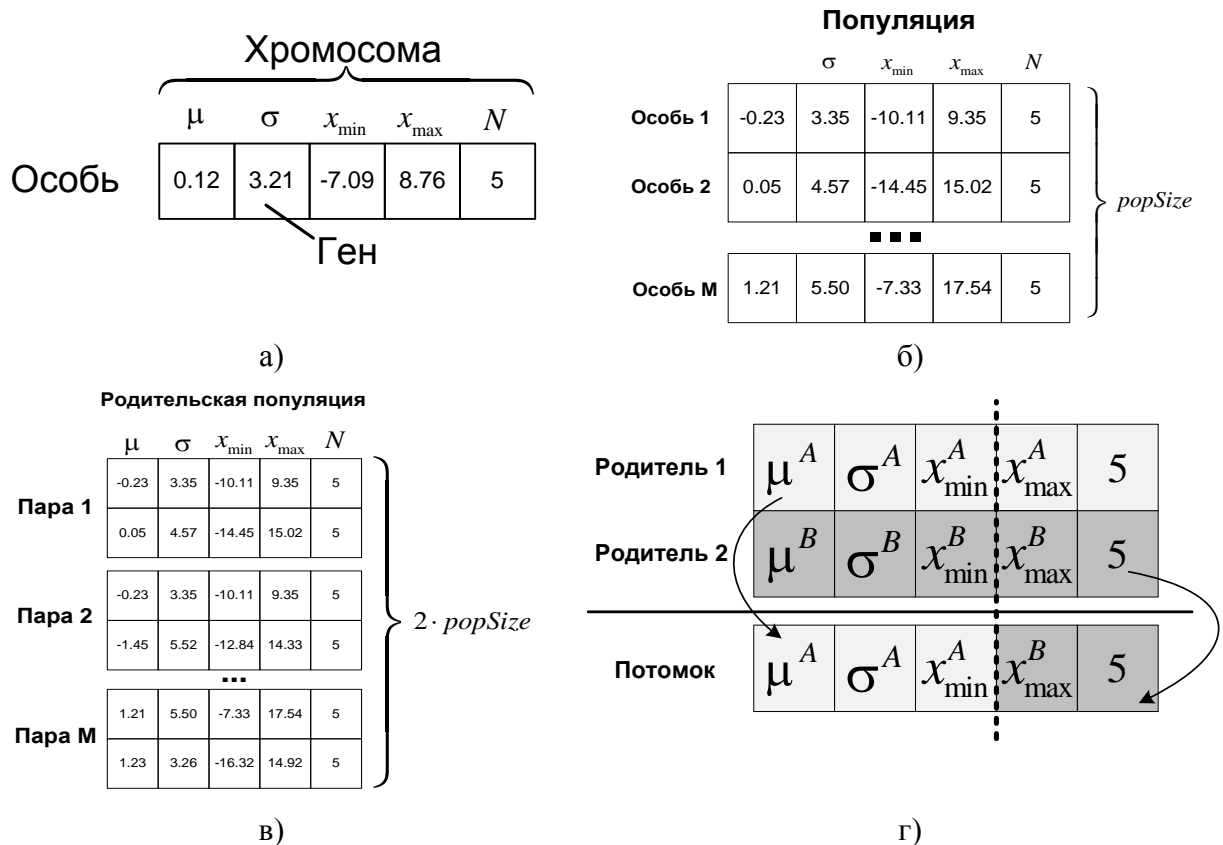


Рисунок 2.2. К установлению соответствия между параметрами распределения и параметрами ГА: а – представление особи, б – пример популяции, в – родительская популяция, полученная в результате селекции, г – пример одноточечного кроссовера

Каждая родительская пара родителя производила одного потомка (рисунок 2.2г). В результате из родительской популяции P' размером $2 \cdot popSize$ особей получалась популяция следующего поколения размером $popSize$.

5. Выполнение **мутации** (этап 6 на рисунок 2.1). **Мутация** – процесс внесения случайных изменений в хромосомы, приводящий к генетическому разнообразию популяции и позволяющий ГА «выбираться» из локальных экстремумов и исследовать более обширное пространство параметров.

Здесь задавалось число особей $mutSize < popSize$, подвергаемых мутации. На каждой итерации ГА после кроссовера случайным образом выбиралось $mutSize$ особей из популяции P . Далее для каждого гена выбранных особей задавались новые значения, выбираемые из генеральной совокупности случайных величин с заданным законом распределения.

6. Проверка выполнения критерия останова. В качестве критерия останова было принят факт появления в популяции особи с функцией приспособленности меньше заданного значения Φ_{min} или выполнения условия

$$\bar{\Phi}(P_i) - \max_i (\Phi(P_i)) < \varepsilon, \quad (2.4)$$

где $\bar{\Phi}(P_i)$ — среднее значение функции приспособленности в популяции, ε — заданная граница колебания функции приспособленности.

Работа алгоритма завершалась, если критерий останова был достигнут, иначе повторялись действия, описанные в пп. 2–5.

Из приведенного выше описания алгоритма использования ГА для оценивания параметров распределения (2.2) по случайной выборке видно, что для решения поставленной задачи требуется обосновать выбор начальных значений следующих параметров ГА:

- границ области изменения параметров-генов;
- размера популяции.
- способа формирования начальной популяции;
- функции приспособленности особи;
- количества особей, подвергаемых процедуре мутации;
- оператора мутации;
- оператор селекции;
- оператора кроссовера.

Для обоснования начальных значений параметров ФР и ПР СПООР $\mu, \sigma, x_{min}, x_{max}$ и областей их поиска потребовалось провести исследование особенностей случайных броуновских блужданий в ограниченной области рассеяния, результаты которого изложены в следующем разделе работы.

2.2 Исследование особенностей одномодальных распределений случайных последовательностей на основе статистического моделирования

2.2.1 Исследование особенностей случайных блужданий броуновской частицы в ограниченной области рассеяния на основе статистического моделирования

Блок-схема алгоритма моделирования случайных блужданий в ограниченной области рассеяния $[x_{\min}, x_{\max}]$ представлена на рисунке 2.3.

В данном алгоритме для каждой частицы ансамбля после каждого шага Монте-Карло проводится сравнение текущей координаты частицы x_n с координатами левой и правой границ. Если координата точки на очередном шаге $x_n > x_{\max}$ то, частицу упруго отражается от стенки и оказывается на расстоянии $x_n - x_{\max}$ слева от точки x_{\max} :

$$x_{\max} - (x_n - x_{\max}) = 2x_{\max} - x_n.$$

Если координата точки на очередном шаге $x_n < x_{\min}$ то, частица упруго отражается от стенки и оказывается на расстоянии $x_{\min} - x_n$ справа от точки x_{\min} :

$$x_{\min} + (x_{\min} - x) = 2x_{\min} - x,$$

и т.д.

После вычисления значений координат на каждом шаге Монте-Карло каждой из частиц ансамбля (траекторий частиц) получают ансамбль траекторий $x_{k,n}$, k – номер частицы, $k = \overline{1, Nm}$, n – номер шага Монте-Карло, вычисляют усредненную по ансамблю реализаций траекторию:

$$\langle x \rangle_n = \frac{1}{Nm} \sum_{k=1}^{Nm} x_{k,n},$$

а также зависимость СКО от номера шага $\langle x^2 \rangle_n$:

$$\langle x^2 \rangle_n = \frac{1}{Nm} \sum_{k=1}^{Nm} x_{k,n}^2.$$

В соответствие с описанным выше алгоритмом были проведено статистическое моделирование случайных блужданий с ограниченной областью рассеяния, цель которых состояла в выявлении функциональных зависимостей между начальной координатой частицы x_0 , координатами левой и правой границ области рассеяния и зависимостями $\langle x \rangle_n, \langle x^2 \rangle_n$, а также их влияние на вид плотности распределения случайного блуждания.

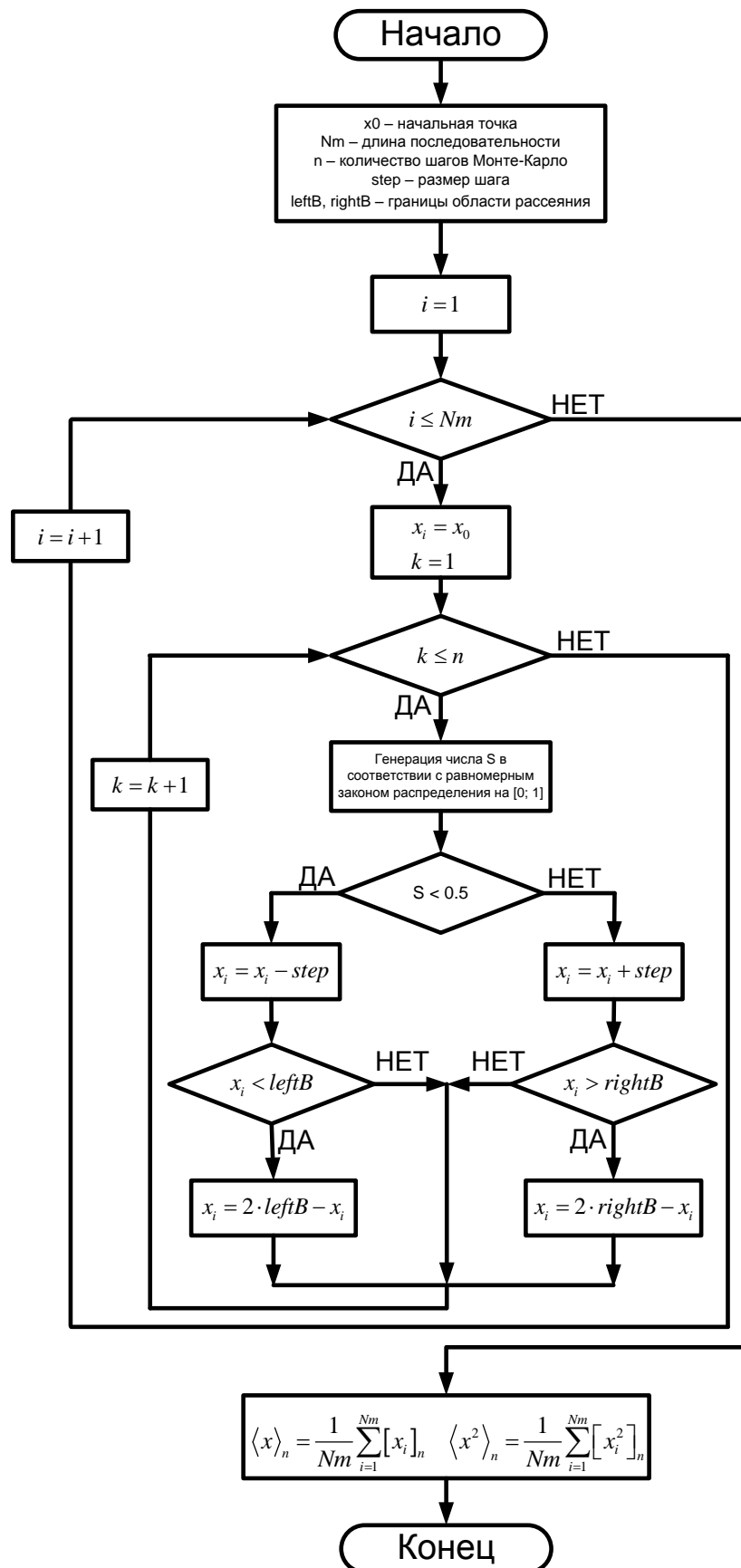
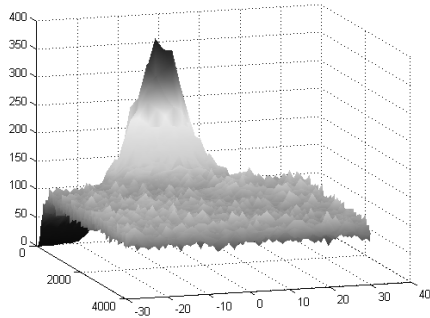
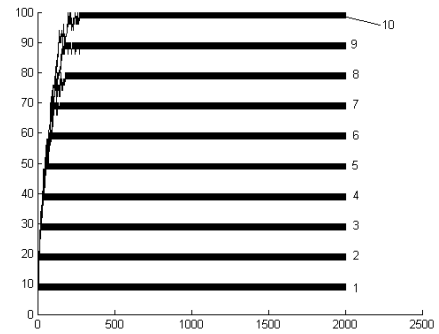


Рисунок 2.3. Блок-схема алгоритма моделирования случайных блужданий в ограниченной области

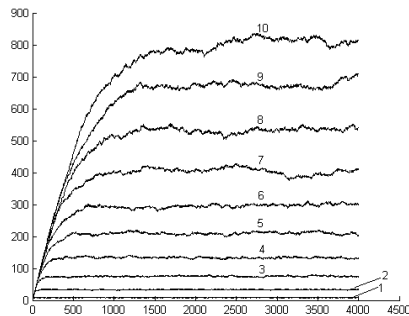
Рассмотрим результаты статистических исследований, представленные на рисунке 2.4.



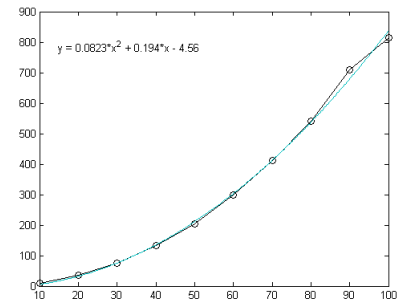
а) Поверхность, задаваемая функцией $Nm \cdot P_n(m)$: $x_0 = 0$, $x_1 = -30$, $x_2 = 30$, $l = 30$, $Nm = 3000$, $n = 50, 60, \dots, 4000$



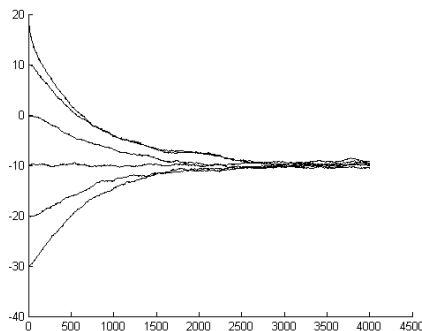
б) Зависимость размаха случайных блужданий от номера шага Монте-Карло: $x_0 = 0$, $x_2 = -x_1$, $|x_1| = |x_2| = l = 10, 20, \dots, 100$, $Nm = 3000$, $n = 1, 2001$



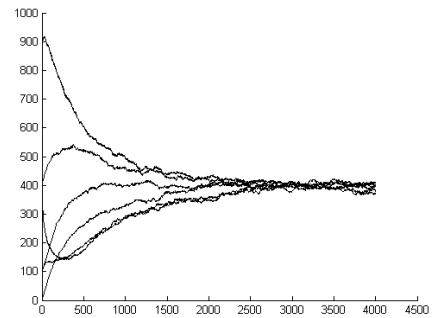
в) Зависимость $\langle x^2 \rangle_n$: $x_0 = 0$, $x_2 = -x_1$, $|x_1| = |x_2| = 2l = 10, 20, \dots, 100$, $Nm = 3000$, $n = 1, 4000$



г) Зависимость $\langle x^2 \rangle_{4000}$ от $2l$: $x_0 = 0$, $Nm = 3000$



д) Зависимость $\langle x \rangle_n$: $x_0 = -30, -20, \dots, 20$, $x_1 = -40$, $x_2 = 30$, $2l = 70$, $Nm = 3000$, $n = 1, 4000$



е) Зависимость $\langle x^2 \rangle_n$: $x_0 = -30, -20, \dots, 20$, $x_1 = -40$, $x_2 = 30$, $2l = 70$, $Nm = 3000$, $n = 1, 4000$

Рисунок 2.4. Результаты статистических исследований случайных блужданий в ограниченной области рассеяния

Из рисунка 2.4 видно, что случайные блуждания в ограниченной области рассеяния имеют следующие свойства:

1. Для случая $x_0 = 0$ и $x_2 = -x_1$ плотность распределения при $n \rightarrow \infty$ стремится к равномерному закону распределения с центром в точке $x_0 = 0$ и областью рассеяния $2l = x_2 - x_1$ (рисунок 2.4а, б).

2. Для случая $x_0 = 0$ и $x_2 = -x_1$ зависимость $\langle x^2 \rangle_n$ при $n \rightarrow \infty$ стремится к некоторому предельному значению, определяемому размером области рассеяния $2l$ (рисунок 2.4в, г).

3. Для случая $|x_1| \neq |x_2|$ плотность распределения при $n \rightarrow \infty$ в не зависимости от координаты начальной точки x_0 стремится к равномерному закону распределения с областью рассеяния $2l = x_2 - x_1$ и средним значением $\bar{x} = x_1 + (x_2 - x_1)/2 = (x_1 + x_2)/2$ (рисунок 2.4д).

4. Для случая $|x_1| \neq |x_2|$ зависимость $\langle x^2 \rangle_n$ при $n \rightarrow \infty$ вне зависимости от координаты x_0 стремится к некоторому предельному значению, определяемому размером области рассеяния $2l$ (рисунок 2.4е).

Результаты проведенных исследований позволяют дать следующие рекомендации для использования на практике математической модели (1.16) в задаче аппроксимации распределения случайной последовательности $x_i, i = 1, M$.

1. Нулевое приближении параметра x_0 нужно выбирать из отрезка $[\tilde{x}_{\min}, \tilde{x}_{\max}]$, где $\tilde{x}_{\min}, \tilde{x}_{\max}$ – минимальное и максимальное значения случайной последовательности $x_i, i = 1, Ns$.

2. Область поиска истинного значения параметра x_0 следует ограничить отрезком $[\tilde{x}_{\min}, \tilde{x}_{\max}]$.

3. Нулевое приближение параметра $\tilde{\sigma}$ нужно оценивать по экспериментальному значению области рассеяния $2l = \tilde{x}_{\max} - \tilde{x}_{\min}$ и зависимости $\langle x^2 \rangle_n - \langle x^2 \rangle_0 = f(2l)$, полученной на основе статистического моделирования (рисунок 2.4г).

4. Область поиска истинного значения параметра σ следует ограничить отрезком $[0, \tilde{\sigma}]$.

2.2.2 Аналитическое исследование особенностей случайных блужданий броуновской частицы в ограниченной области рассеяния

Для подтверждения результатов моделирования, описанных в предыдущем разделе, получено аналитическое решение задачи случайных блужданий в ограниченной области рассеяния.

2.2.2.1 Свойства случайных блужданий в неограниченной области рассеяния

Основные формулы, описывающие вероятности нахождения броуновской частицы в заданной точке приведены в разделе 1.1.1.1 (формулы (1.4)—(1.6))

Вычислим зависимость координат усредненной по ансамблю реализаций частиц, совершающих независимые перемещения, от номера шага перемещения n

$$\langle x \rangle_n = \frac{1}{Nm} \sum_{i=1}^{Nm} [x_i]_n,$$

где $i = \overline{1, Nm}$, Nm – число частиц в ансамбле, для случая $x_0 = 0$ описывается выражением (2.5) [69]

$$\langle x \rangle = (p - q)n, \quad (2.5)$$

для случая $x_0 \neq 0$ описывается выражением (2.6)

$$\langle x \rangle = x_0 + (p - q)n. \quad (2.6)$$

Зависимость квадрата смещения частицы, усредненной по ансамблю частиц, совершающих независимые перемещения, от номера шага перемещения n

$$\langle x^2 \rangle_n = \frac{1}{Nm} \sum_{i=1}^{Nm} [x_i^2]_n,$$

где $i = \overline{1, Nm}$, Nm – число частиц в ансамбле, для случая $x_0 = 0$ описывается выражением (2.7) [69]

$$\langle x^2 \rangle_n = 4pqn, \quad (2.7)$$

для случая $x_0 \neq 0$ описывается выражением (2.8)

$$\langle x^2 \rangle_n = x_0^2 + 4pqn, \quad (2.8)$$

Если $p = q = \frac{1}{2}$, выражения (2.5)–(2.8), соответственно, принимают следующий вид:

$$\langle x \rangle_n = 0, \quad (2.9)$$

$$\langle x \rangle_n = x_0, \quad (2.10)$$

$$\langle x^2 \rangle = n, \quad (2.11)$$

$$\langle x^2 \rangle = x_0^2 + n. \quad (2.12)$$

Из (2.7), (2.8), (2.11), (2.12) видно, что дисперсия случайного блуждания зависит линейно от числа шагов частицы, совершающей случайные блуждания, что соответствует результату, полученному А.

Эйнштейном для броуновского движения, который показал [83], что дисперсия линейно зависит от времени t , которое в данном случае эквивалентно числу случайных шагов n .

2.2.2.2 Свойства случайных блужданий в области, ограниченной с одной стороны

Рассмотрим, следуя [23], свойства случайных блужданий в области, ограниченной справа абсолютно упругой отражающей стенкой, расположенной в точке x_{\max} . Присутствие стенки обеспечивает с вероятностью, равной единице, что при следующем толчке частица будет двигаться в направлении, противоположном направлению движения на предыдущем шаге. Изобразим положение частицы на плоскости (x, t) (рисунок 2.5).

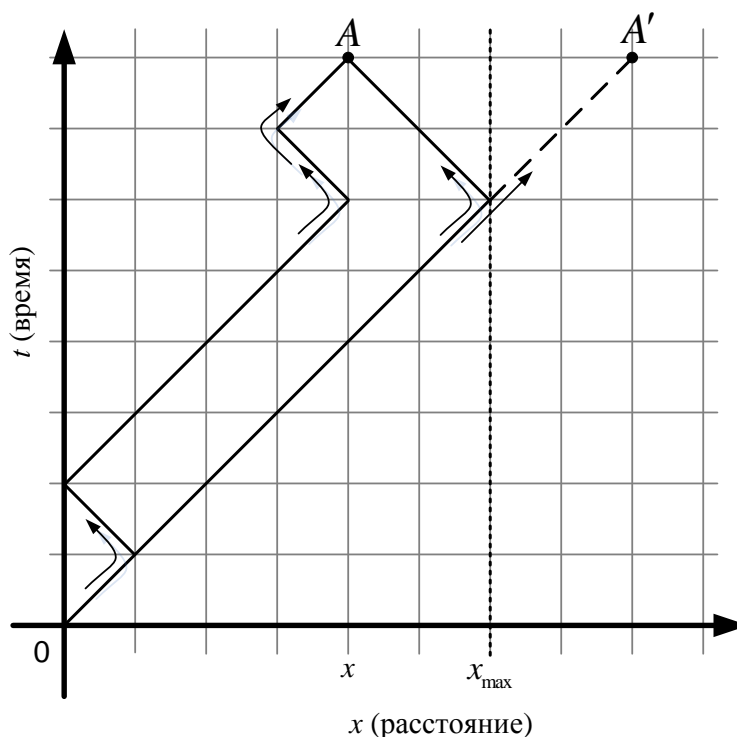


Рисунок 2.5. Случайное блуждание частицы с одной отражающей стенкой

Так как при каждом толчке частица перемещается на одну единицу «вверх» и с вероятностью q на единицу вправо или с вероятностью p влево, то траектория частицы на плоскости (x, t) представляет собой ломаную линию. Если $x = x_{\max}$, то при очередном толчке частица сдвинется на единицу влево и окажется в точке $x = x_{\max} - 1$.

Для подсчета вероятностей $P(x = m)$ мысленно откинем стенку и разрешим частице двигаться свободно. Идеализированные пути, приводящие частицу в симметрично расположенные относительно стенки $x = x_{\max}$ точки A и A' , представлены на рисунке 2.5. Из рисунка 2.5 видно, что для попадания в действительности частицы в точку A , необходимо и достаточно, чтобы при

движении в области без отражающей стенки частица достигла либо точки A , либо точки A' . Но вероятность попасть в точку A в ситуации без отражающей стенки вычисляется для случая $q \neq p$ в соответствие с (1.4), а для случая

$q = p = \frac{1}{2}$ в соответствие с (1.5).

Аналогично, вероятность попасть в точку A' с абсциссой $2x_{\max} - m$ равна

$$P_n(m, x_{\max}) = P_n(x = m) + P_n(x = 2x_{\max} - m). \quad (2.13)$$

При больших значениях n при $q = p = \frac{1}{2}$ выражение (2.13) в соответствии с локальной теоремой Муавра-Лапласа аппроксимируется следующим выражением:

$$P\{x = m\} \approx \frac{1}{\sqrt{2\pi n}} e^{\frac{-m^2}{2n}} + \frac{1}{\sqrt{2\pi n}} e^{\frac{-(2x_{\max} - m)^2}{2n}}. \quad (2.14)$$

Из (2.14) видно, что в рассматриваемом случае вероятность представляет сумму двух нормальных распределений одного расположенного в точке $x = m$, второго – в точке $x = 2x_{\max} - m$. Соотношение (2.13) с физической точки зрения может быть истолковано следующим образом [20]. Предположим, что в начальный момент времени в точках $x = 0$, $x = 2x_{\max} - m$ имелось большое количество частиц (ансамбль частиц). Далее частицы начинают перемещаться под действием случайных толчков независимо друг от друга. Тогда, $P_n(x = m)$ есть доля частиц, оказавшихся на расстоянии m от начала отсчета, $P_n(x = 2x_{\max} - m)$ – доля частиц, оказавшихся на расстоянии m от точки x_{\max} .

Для случая $x_0 \neq 0$, $q \neq p$ (2.14) принимает следующий вид:

$$P_n(m) = \frac{n!}{\left(\frac{n + (m - x_0)}{2}\right)! \left(\frac{n - (m - x_0)}{2}\right)!} p^{\frac{m+n}{2}} q^{\frac{n-m}{2}} + \frac{n!}{\left(\frac{n + (2x_{\max} - (m - x_0))}{2}\right)! \left(\frac{n - (2x_{\max} - (m - x_0))}{2}\right)!} p^{\frac{m+n}{2}} q^{\frac{n-m}{2}}, \quad (2.15)$$

для случая $q = p = \frac{1}{2}$:

$$P_n(m) = C_n^{\frac{n+m-x_0}{2}} \left(\frac{1}{2}\right)^n + C_n^{\frac{n-(2x_{\max}-(m-x_0))}{2}} \left(\frac{1}{2}\right)^n. \quad (2.16)$$

2.2.2.3 Свойства случайных блужданий в ограниченной области рассеяния

Добавление дополнительной отражающей стенки в точке $x = x_{\min}$ приведет к появлению бесконечной системы, состоящей из источников частиц (1.2), координаты которых вычисляются в соответствие с (1.13), поэтому вероятность нахождения частицы в точке с координатой $x = m$ можно вычислить по следующей формуле

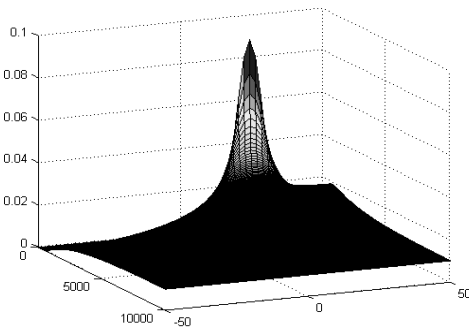
$$P_n(m; x_0, x_{\min}, x_{\max}) = P_n(x = m) + \sum_{g=0}^{\infty} P_{2g+1}^{\pm}(x = m - x_0 - x_{2g+1}^{\pm}) + \sum_{g=1}^{\infty} P_{2g}^{\pm}(x = m - x_0 - x_{2g}^{\pm}), \quad (2.17)$$

где $P_n(x = m)$, $\sum_{g=0}^{\infty} P_{2g+1}^{\pm}(x = m - x_0 - x_{2g+1}^{\pm})$, $\sum_{g=1}^{\infty} P_{2g}^{\pm}(x = m - x_0 - x_{2g}^{\pm})$ вычисляются в соответствии с (1.4), (1.5) а координаты x_{2g+1}^{\pm} , x_{2g}^{\pm} в соответствии с (1.13), а начало системы координат находится в середине области рассеяния (локальная система координат) (см. рисунок 1.2). В этой связи, при задании значений координат отражающих границ и координаты центра рассеяния в системе координат с центром в точке $x = 0$ (абсолютная система координат) нужно пересчитывать координаты $x_0, x_{\min}, x_{\max}, m$ из абсолютной в локальную систему координат по следующим очевидным формулам:

$$x_0^{local} = x_0 - dx, \quad x_{\min}^{local} = -l, \quad x_{\max}^{local} = l, \quad m^{local} = m - dx,$$

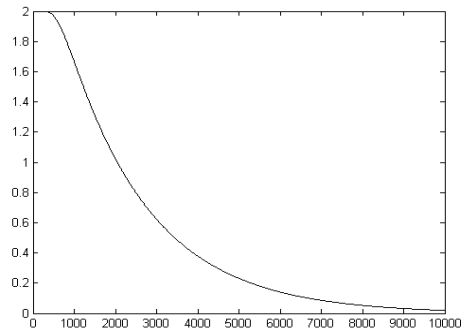
$$\text{где } l = \frac{x_{\max} - x_{\min}}{2}, \quad dx = \frac{x_{\max} + x_{\min}}{2}.$$

Рассмотрим результаты расчетов значений ФР, а также зависимостей $\langle x \rangle, \langle x^2 \rangle$ для различных значений x_0, x_{\min}, x_{\max} , представленные на рисунке 2.6.



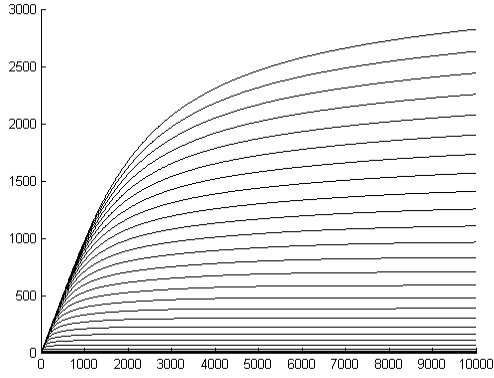
а) Поверхность, задаваемая функцией $Nm \cdot P_n(m)$: $x_0 = 0$, $x_{\min} = -50$, $x_{\max} = 50$,

$$l = 50, \quad m = x_{\min}, x_{\min} + 2, \dots, x_{\max}, \\ n = 50, 60, \dots, 10000$$

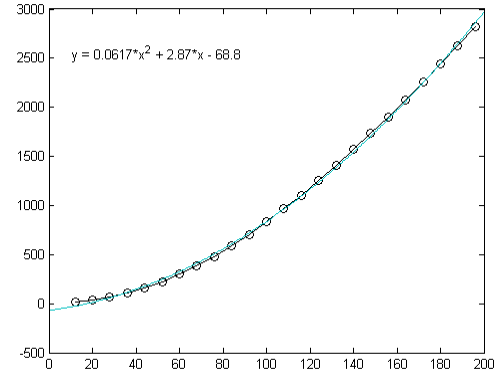


б) Зависимость $\frac{\max(P_n(m)) - \min(P_n(m))}{\max(P_n(m)) + \min(P_n(m))}$ от $n = \overline{1, 10000}$

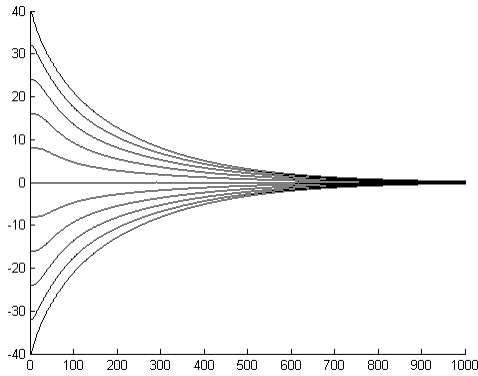
Рисунок 2.6. Результаты исследования случайных блужданий с ограниченной областью рассеяния (начало)



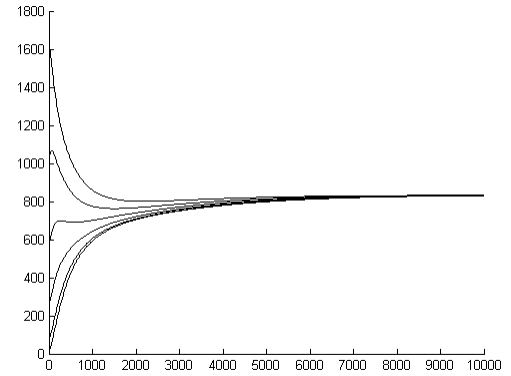
в) Зависимость $\langle x^2 \rangle_n$: $x_0 = 0$,
 $|x_{\min}| = |x_{\max}| = l$, $l = 6, 14, \dots, 200$ $n = \overline{1, 10000}$



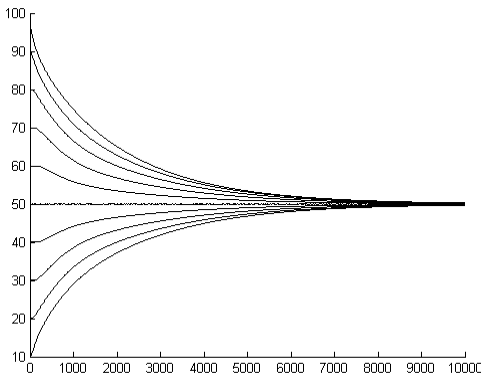
г) Зависимость $\langle x^2 \rangle_n$ от $2l$



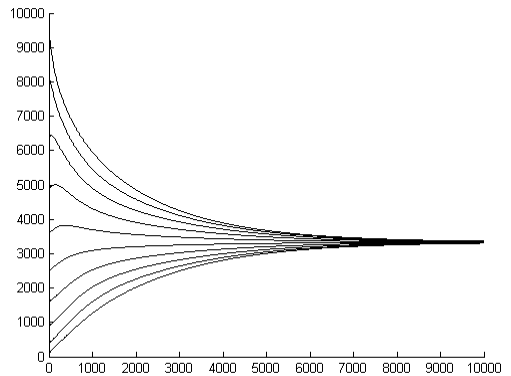
д) Зависимость $\langle x \rangle_n$: $x_0 = -40, -30, \dots, 40$
 $x_{\min} = -x_{\max}$, $|x_1| = |x_2| = l/2 = 50$, $n = \overline{1, 10000}$



е) Зависимость $\langle x^2 \rangle_n$:
 $x_0 = -40, -30, \dots, 40$ $n = \overline{1, 10000}$



ж) Зависимость $\langle x \rangle_n$: $x_0 = 10, 20, \dots, 100$
 $x_{\min} = 0$, $x_{\max} = 100$, $l/2 = 50$, $n = \overline{1, 10000}$



з) Зависимость $\langle x^2 \rangle_n$: $x_0 = 10, 20, \dots, 100$
 $x_{\min} = 0$, $x_{\max} = 100$, $l/2 = 50$, $n = \overline{1, 10000}$

Рисунок 2.6 (окончание)

Из рисунка 2.6 видно, что случайные блуждания в ограниченной области рассеяния имеют следующие свойства:

1. Для случая $x_0 = 0$ и $x_2 = -x_1$ плотность распределения при $n \rightarrow \infty$ стремится к равномерному закону распределения с центром в точке $x_0 = 0$ и областью рассеяния $2l = x_2 - x_1$ (рисунок 2.6а, б).

2. Для случая $x_0 = 0$ и $x_2 = -x_1$ зависимость $\langle x^2 \rangle_n$ при $n \rightarrow \infty$ стремится к некоторому предельному значению, определяемому размером области рассеяния $2l$ (рисунок 2.6в, г).

3. Зависимость $\langle x \rangle_n$ при $n \rightarrow \infty$ не зависит от координаты начальной точки x_0 и стремится к равномерному закону распределения с областью рассеяния $2l = x_2 - x_1$ и средним значением $\bar{x} = x_1 + (x_2 - x_1)/2 = (x_1 + x_2)/2$ (рисунок 2.6д, ж).

4. Зависимость $\langle x^2 \rangle_n$ при $n \rightarrow \infty$ не зависит от координаты x_0 и стремится к некоторому предельному значению, определяемому размером области рассеяния $2l$ (рисунок 2.6е, з).

5. Для нахождения предельного значения зависимость $\langle x^2 \rangle_n$ при $n \rightarrow \infty$ следует использовать зависимость, представленную на рисунке 2.6г, которая аппроксимируется с удовлетворительной точностью следующей функцией:

$$f(2l) = 0.0617(2l)^2 + 2.87(2l) - 68.8. \quad (2.18)$$

Таким образом, полученные ранее результаты на основе статистического моделирования согласуются с аналитическими выводами.

2.2.3 Выводы

Существуют принципиальные отличия между случайными блужданиями с неограниченной областью рассеяния, у которых средний квадрат смещения броуновской частицы, линейно неограниченно увеличивается с течением времени, и ограниченной областью рассеяния, у которых данная величина стремится к некоторому предельному значению, зависящему от квадрата размера области случайных блужданий и независящему от точки x_0 .

Результаты проведенных исследований позволяют дать следующие рекомендации по выбору следующих параметров ГА, используемых в задаче аппроксимации ФР и ПР СПООР $x_i, i = 1, Ns$.

1. В качестве нулевого приближения параметра μ следует использовать произвольное число из отрезка $[\tilde{x}_{\min}, \tilde{x}_{\max}]$, где $\tilde{x}_{\min}, \tilde{x}_{\max}$ – минимальное и максимальное значения случайной последовательности $x_i, i = 1, Ns$.

2. Область поиска истинного значения параметра μ следует ограничить отрезком $[\tilde{x}_{\min}, \tilde{x}_{\max}]$.

3. Нулевое приближение параметра $\tilde{\sigma}$ нужно оценивать по экспериментальному значению области рассеяния $2l = \tilde{x}_{\max} - \tilde{x}_{\min}$ и зависимости $\langle x^2 \rangle_n = f(l)$, значение которой вычисляются в соответствии с (2.18).

4. Область поиска истинного значения параметра σ следует ограничить отрезком $[0, \tilde{\sigma}]$.

2.3 Анализ точности оценок параметров одномодальных распределений случайных последовательностей с ограниченной областью рассеяния с помощью ГА

В ходе проведенных исследований были изучены СПООР со следующими параметрами: $N=5$, $\sigma=10$, $x_{\min}=-l$, $x_{\max}=l$, где $2l$ – размер области рассеяния, для 4 различных значений интервала рассеяния – $2l = 2\sigma, 3\sigma, 5\sigma, 10\sigma$ ($numL=4$) и 3 различных положений центра распределения (1.16) относительно границ интервала: $\mu = 0, l/2, l$ ($numMu=3$). Выбор данного набора параметров обусловлен тем, что он позволяет охватить диапазон всевозможных характеристических вариантов расположения центра распределения и отношения размера области рассеяния к СКО распределения. Таким образом, было изучено

$$numSeq = numMu \times numL = 4 \times 3 = 12$$

различных распределений нормальных случайных последовательностей с ограниченной областью рассеяния.

В качестве эталона для функции приспособленности в рассматриваемой задаче использовались выборка X значений изучаемой случайной величины. Блок-схема алгоритма вычисления ФП для особи $P_i = \{\mu_i, \sigma_i, x_{\min i}, x_{\max i}, 5\}$ представлена на рисунке 2.7.

В проведенных исследованиях использовались следующие настройки ГА:

- Селекция: $S = \langle \text{равномерная; турнирная; на основе рулетки} \rangle$;
- Мутация: $M = \langle \text{адаптивная} \rangle$ (т.к. в задаче присутствуют ограничения)
- Кроссовер: $C = \langle \text{одноточечный; двухточечный; усредненный; разбросанный, эвристический} \rangle$;
- Доля кроссовера: $Cfr = \langle 0,3; 0,6; 0,9 \rangle$;
- Доля мутации $1 - Cfr$;
- Размер популяции: $P = \langle 5, 10, 25 \rangle$.

Таким образом, каждый эксперимент описывался кортежем:

$$\langle S_s, M_m, C_c, Cfr_f, P_p \rangle,$$

поэтому всего было рассмотрено

$$numT = numS \times numM \times numC \times numCfr \times numP = 3 \times 1 \times 5 \times 3 \times 3 = 135$$

различных комбинаций настроек.

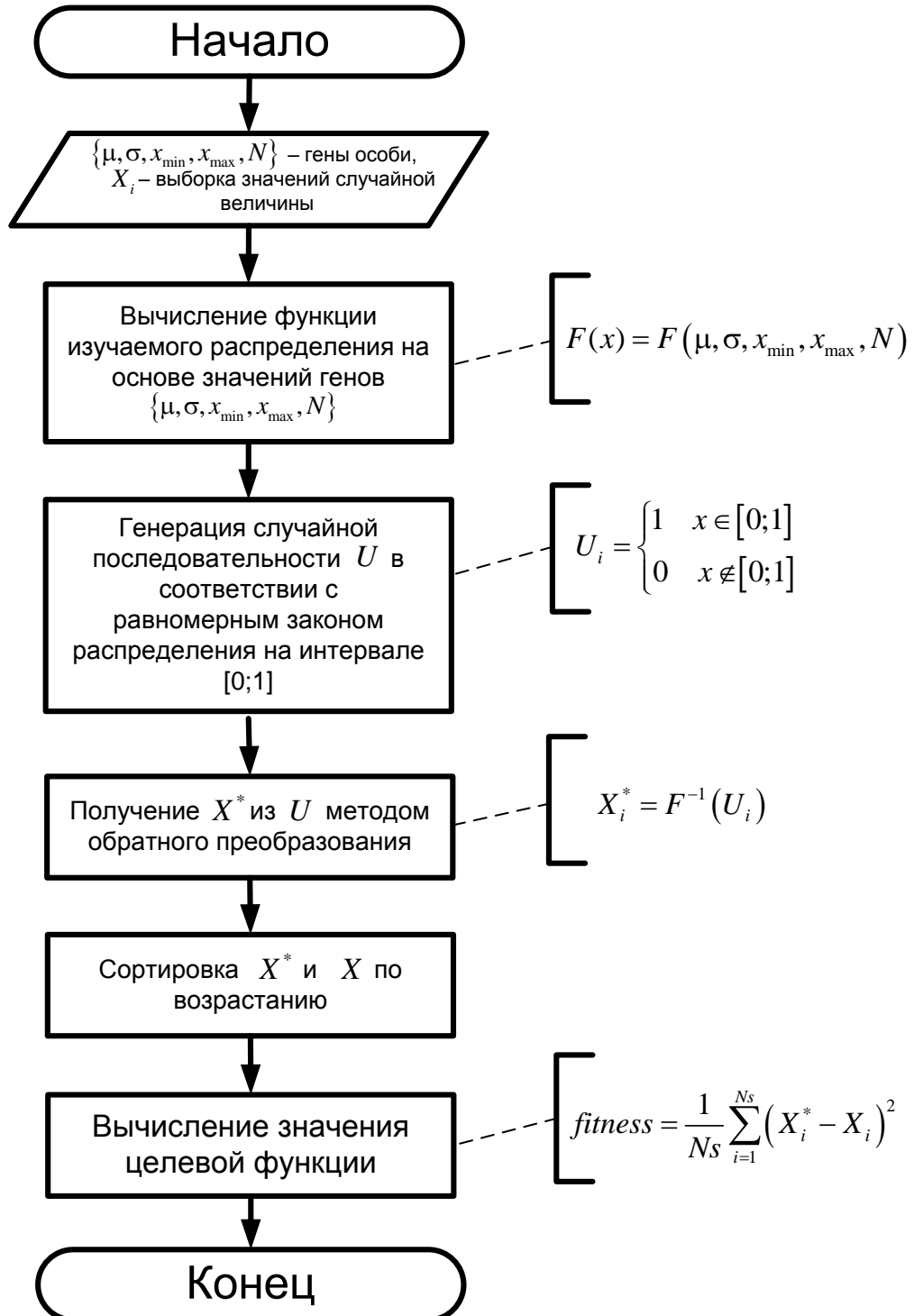


Рисунок 2.7. Блок схема алгоритма вычисления функции приспособленности: X - анализируемая случайная последовательность

Для каждого набора настроек $p = \overline{1, numT}$ было выполнено 10 независимых статистических испытаний ($numExp = 10$), в ходе которых на каждом k -ом шаге:

1) генерировалась случайная последовательность $\{x_i^*\}^{(k,p)}$, $i = \overline{1, Ns}$, ($Ns = 500$), $k = \overline{1, numExp}$ в соответствие с ФР $F(\tilde{\mu}^{\{k,p\}}, \tilde{\sigma}^{\{k,p\}}, \tilde{x}_{\min}^{\{k,p\}}, \tilde{x}_{\max}^{\{k,p\}}, 5)$;

2) случайная последовательность $\{x_i^*\}^{\{k,p\}}$ упорядочивалась по возрастанию: $X_i^{\{k,p\}} = sort(x_i^{\{k,p\}})$;

3) вычислялась значения ФП: $fitness^{\{k,p\}} = \frac{1}{Ns} \sum_{i=1}^{Ns} (X_i^{\{k,p\}} - X_i)^2$.

После проведения $numExp$ испытаний из ансамбля хромосом $\tilde{\mu}^{\{k,p\}}, \tilde{\sigma}^{\{k,p\}}, \tilde{x}_{\min}^{\{k,p\}}, \tilde{x}_{\max}^{\{k,p\}}, 5$ выбиралась m -ая хромосома, у которой значений ФП $fitness^{\{m,p\}}$ было наименьшим.

Далее значение $fitness^{\{m,p\}}$ сравнивалось с выбранным критерием останова (см. рис. 2.1), при выполнении которого выполнение ГА для данного набора параметров прекращалось.

Таким образом, общее число запусков ГА составило

$$numTotal = numT \times numSeq \times numExp = 135 \times 12 \times 10 = 16200.$$

Затем для каждого набора настроек $p = \overline{1, numT}$ вычислялась оценка величины, характеризующей отличие между теоретической и экспериментальной ФР:

$$[\Delta_{integr}]^{(p)} = \frac{\sum_i^{Ns_p} \left(F_{teor}(x_i, \mu, \sigma, x_{\min}, x_{\max}, 5) - \left[F_{pract}(x_i, \tilde{\mu}^{\{m,p\}}, \tilde{\sigma}^{\{m,p\}}, \tilde{x}_{\min}^{\{m,p\}}, \tilde{x}_{\max}^{\{m,p\}}, 5) \right] \right)^2}{\sum_i^{Ns_p} F_{teor}(x_i, \mu, \sigma, x_{\min}, x_{\max}, 5)^2}.$$

Далее из ансамбля $[\Delta_{integr}]^{(p)}$ выбирались те наборы параметров, у которых значение данной величины было наименьшим.

Наилучшие значения наборов параметров, а также соответствующие значения $[\Delta_{integr}]$ представлены в таблице 2.1.

Таблица 2.1. Результаты оценивания параметров нормального распределения с ограниченной областью рассеяния

| № набора параметров | Параметры | μ | σ | x_{\min} | \tilde{x}_{\max} | Δ_{integr} |
|---------------------|------------------------|--------|----------|------------|--------------------|--------------------------|
| 1 | Теоретические значения | 0.000 | 10.000 | -10.000 | 10.000 | $2.692 \cdot 10^{-2}$ |
| | Наилучший результат | -1.473 | 7.968 | -9.730 | 9.990 | |
| 2 | Теоретические значения | 5.000 | 10.000 | -10.000 | 10.000 | $3.920 \cdot 10^{-3}$ |
| | Наилучший результат | 3.486 | 7.788 | -9.910 | 9.870 | |
| 3 | Теоретические значения | 10.000 | 10.000 | -10.000 | 10.000 | $4.000 \cdot 10^{-5}$ |
| | Наилучший результат | 5.166 | 8.547 | -9.790 | 9.990 | |
| 4 | Теоретические значения | 0.000 | 10.000 | -15.000 | 15.000 | $1.590 \cdot 10^{-3}$ |
| | Наилучший результат | -0.371 | 9.481 | -14.925 | 14.685 | |
| 5 | Теоретические значения | 7.500 | 10.000 | -15.000 | 15.000 | $4.600 \cdot 10^{-4}$ |
| | Наилучший результат | 7.144 | 8.693 | -9.915 | 14.955 | |
| 6 | Теоретические значения | 15.000 | 10.000 | -15.000 | 15.000 | $1.000 \cdot 10^{-5}$ |
| | Наилучший результат | 11.281 | 9.258 | -14.775 | 14.985 | |
| 7 | Теоретические значения | 0.000 | 10.000 | -25.000 | 25.000 | $3.320 \cdot 10^{-3}$ |
| | Наилучший результат | 0.553 | 10.253 | -20.775 | 24.775 | |
| 8 | Теоретические значения | 12.500 | 10.000 | -25.000 | 25.000 | $9.000 \cdot 10^{-5}$ |
| | Наилучший результат | 12.370 | 9.945 | -14.025 | 24.675 | |
| 9 | Теоретические значения | 25.000 | 10.000 | -25.000 | 25.000 | $9.000 \cdot 10^{-5}$ |
| | Наилучший результат | 20.186 | 8.562 | -2.625 | 24.875 | |
| 10 | Теоретические значения | 0.000 | 10.000 | -50.000 | 50.000 | $6.000 \cdot 10^{-4}$ |
| | Наилучший результат | 0.072 | 9.658 | -25.650 | 25.450 | |
| 11 | Теоретические значения | 25.000 | 10.000 | -50.000 | 50.000 | $4.590 \cdot 10^{-3}$ |
| | Наилучший результат | 24.647 | 10.339 | -0.650 | 45.350 | |
| 12 | Теоретические значения | 50.000 | 10.000 | -50.000 | 50.000 | $1.000 \cdot 10^{-4}$ |
| | Наилучший результат | 47.359 | 9.736 | 13.850 | 49.950 | |

Соответствующие значения параметров ГА (селекция, мутация, кроссовер, доля кроссовера, размер популяции) представлены в таблице 2.2.

Таблица 2.2. Настройки ГА для каждого набора параметров

| № | Селекция | Мутация | Кроссовер | Доля кроссовера | Размер популяции |
|----|-------------|------------|---------------|-----------------|------------------|
| 1 | равномерная | адаптивная | одноточечный | 0.9 | 25 |
| 2 | Рулетка | адаптивная | двухточечный | 0.6 | 25 |
| 3 | равномерная | адаптивная | одноточечный | 0.3 | 25 |
| 4 | равномерная | адаптивная | усредненный | 0.6 | 25 |
| 5 | Турнирная | адаптивная | разбросанный | 0.6 | 25 |
| 6 | Турнирная | адаптивная | эвристический | 0.9 | 25 |
| 7 | равномерная | адаптивная | эвристический | 0.3 | 25 |
| 8 | Турнирная | адаптивная | эвристический | 0.3 | 25 |
| 9 | равномерная | адаптивная | одноточечный | 0.3 | 25 |
| 10 | равномерная | адаптивная | одноточечный | 0.3 | 25 |
| 11 | равномерная | адаптивная | эвристический | 0.6 | 25 |
| 12 | Турнирная | адаптивная | эвристический | 0.6 | 25 |

Гистограммы случайных последовательностей, теоретические функции плотности распределения и наилучшие аппроксимированные функции плотности распределения, соответствующие наборам параметров, приведенным в таблице 2.1, представлены на рисунке 2.8.

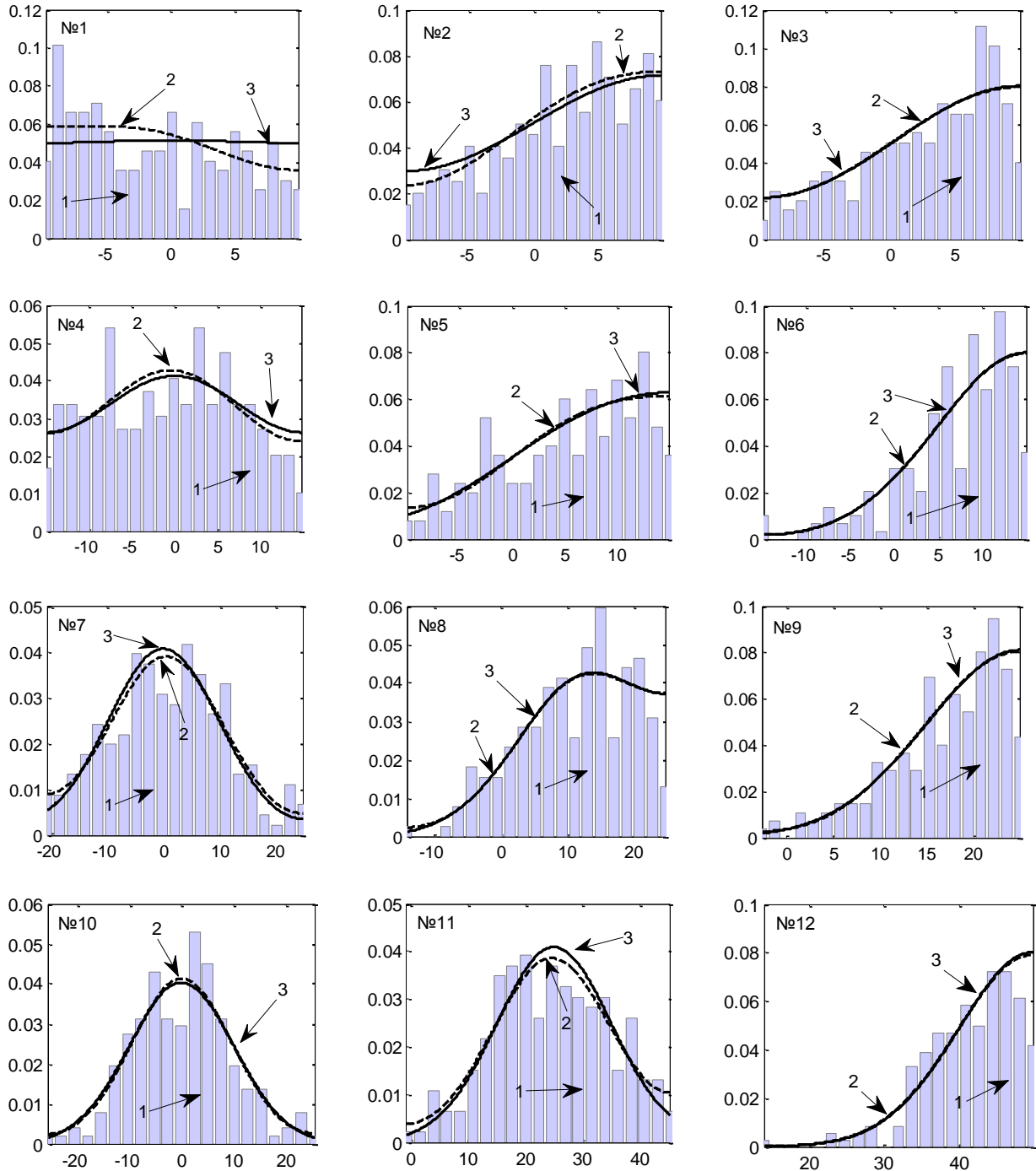


Рисунок 2.8. Функции плотности распределения (1.16), при вычислении которых использованы значения параметров, представленные в 2.1 и настройки ГА, представленные в 2.2: 1 – гистограмма случайной последовательности, 2 – экспериментальная функция плотности распределения, 3 – теоретическая функция плотности распределения

Из таблиц 2.1, 2.2 и рисунка 2.8 видно, что наилучшие результаты получаются при следующих настройках ГА:

- селекция: **равномерная** или **турнирная**;
- мутация: **адаптивная**;
- кроссовер – **одноточечный** или **эвристический**;
- доля кроссовера находится в интервале от **0.2** до **0.7**;
- размер популяции: **25–50**.

Отметим, что увеличение размера популяции, увеличивает разнообразие входящих в нее особей и, соответственно, увеличивает вероятность нахождения оптимального решения. Однако увеличение размера популяции пропорционально увеличивает объем, и, соответственно, время вычислений. В этой связи размер популяции приходится для каждой конкретной задачи выбирать эмпирически, исходя из соотношения между временем расчетов и точностью оценок параметров плотности нормального распределения с ограниченной областью рассеяния случайных последовательностей.

2.4 Методика оценивания параметров одномодальных распределений случайных последовательностей с ограниченной областью рассеяния, основанный на использовании ГА

Таким образом, основываясь на результатах, полученных ранее в данной главе, можно предложить следующую методику оценки параметров ФР СПООР $x_i, i = \overline{1, N_s}$, которая реализуется выполнением следующей последовательности действий:

1. Вычисление оценки координаты левой границы области рассеяния распределения $\tilde{x}_{\min} = \min_i(x_i)$.
2. Вычисление оценки координаты правой границы области рассеяния распределения $\tilde{x}_{\max} = \max_i(x_i)$.
3. Задание количества пар фиктивных источников $N = 5$.
4. Задание интервала ограничения для параметра $\mu: [\tilde{x}_{\min}, \tilde{x}_{\max}]$.
5. Задание интервала ограничения для параметра $\sigma: [0, \sigma_{\max}]$, где σ_{\max} вычисляется по зависимости (2.18).
6. Задание настроек ГА:
 - селекция: равномерная или турнирная;
 - мутация: адаптивная;
 - кроссовер: одноточечный или эвристический;
 - доля кроссовера: в интервале от 0.2 до 0.7;
 - размер популяции: 25–50.

7. Вычисление, в соответствии с ГА, значения оценок параметров μ, σ .

Программная библиотека, содержащая функции, необходимые для реализации описанного алгоритма в MATLAB, представлена в Приложении В.

2.5 Выводы по главе

Анализ результатов применения ГА в задаче нахождения параметров случайных последовательностей с нормальным распределением с ограниченной областью рассеяния позволяет сделать следующие выводы:

1. Предложен алгоритм нахождения параметров данного распределения, основанный на использовании ГА.

2. Обоснованы рекомендации выбора областей поиска значений параметров ФП СПООР и настроек ГА.

3. Получены оценки точности параметров ФР СПООР и интегрального показателя, характеризующего качество оценивания изучаемой плотности распределения в целом, подтверждающие работоспособность предложенного алгоритма.

4. Обоснован выбор настроек ГА, обеспечивающих наилучшую точность оцениваемых параметров нормального распределения с ограниченной областью рассеяния.

Результаты исследования, приведенного в данной главе, опубликованы в [60,63,64].

Equation Chapter (Next) Section 1

3 Разработка математического и алгоритмического обеспечения для оценивания параметров двумодальных распределений случайных последовательностей с ООР

В данной главе рассматриваются следующие вопросы:

- постановка задачи оценивания параметров двумодальных распределений случайных последовательностей с помощью ГА;
- обоснование рекомендаций по выбору настроек ГА при оценивании параметров двумодальных нормальных распределений случайных величин с ограниченной областью рассеяния;
- анализ точности оценок параметров одномодальных распределений случайных последовательностей с ограниченной областью рассеяния с помощью ГА;
- обоснование математического и алгоритмического обеспечения для оценивания параметров двумодальных нормальных распределений случайных величин с ограниченной областью рассеяния на основе ГА.

3.1 Постановка задачи оценивания параметров двумодальных распределений случайных последовательностей с помощью ГА

Основные свойства двумодальных ФР и ПР СПООР были описаны в разделе 1.1.2.2. Напомним, что двумодальные ФР и ПР СПООР вычисляются по формулам:

$$\begin{aligned} F_{2\text{mod}}(x) &= F_{LAD}(x, \mu_1, \sigma_1, a_1, b_1) \cdot \alpha + F_{LAD}(x, \mu_2, \sigma_2, a_2, b_2) \cdot (1 - \alpha), \\ f_{2\text{mod}}(x) &= f_{LAD}(x, \mu_1, \sigma_1, a_1, b_1) \cdot \alpha + f_{LAD}(x, \mu_2, \sigma_2, a_2, b_2) \cdot (1 - \alpha). \end{aligned} \quad (3.1)$$

где

- μ_1 – математическое ожидание первой составляющей;
- σ_1 – математическое ожидание первой составляющей;
- a_1, b_1 – границы области рассеяния первой составляющей;
- μ_2 – математическое ожидание первой составляющей;
- σ_2 – математическое ожидание первой составляющей;
- a_2, b_2 – границы области рассеяния первой составляющей;
- α – доля первой составляющей в общем распределении,

$$f_{LAD}(x; x_0, \sigma, l) = A \left[\varphi(x; x_0, \sigma, l) + \sum_{g=0}^{\infty} \varphi_{2g+1}^{\pm}(x; x_0, \sigma, l) + \sum_{g=1}^{\infty} \varphi_{2g}^{\pm}(x; x_0, \sigma, l) \right],$$

где A – нормировочный коэффициент, определяемый из условия

$$\int_a^b f_{LAD}(\xi; x_0, \sigma, l) d\xi = 1,$$

$$\varphi(x; x_0, \sigma, l) = \exp\left[-(x - x_0)^2 / 2\sigma^2\right],$$

$$\varphi_{2g+1}^{\pm}(x; x_0, \sigma, l) = \exp\left[-(x - x_{2g+1}^{\pm})^2 / 2\sigma^2\right],$$

$$\varphi_{2g}^{\pm}(x; x_0, \sigma, l) = \exp\left[-(x - x_{2g}^{\pm})^2 / 2\sigma^2\right],$$

$$F_{LAD}(x; x_0, \sigma, l) = \int_a^x f_{LAD}(\xi; x_0, \sigma, l) d\xi,$$

здесь $x_{2g+1}^{\pm}, x_{2g}^{\pm}$ вычисляются по формулам:

$$x_{2g}^{\pm} = \pm 4gl + x_0, \quad x_{2g+1}^{\pm} = \pm(4g + 2)l - x_0,$$

где $g = 0, 1, \dots$, по которым можно вычислить координаты любого из мнимых источников.

Из (3.1) видно, что изучаемое распределение является, вообще говоря, 10-ти параметрическим. Однако, как было показано в разделе 1.1.1.3 на примере одномодального распределения, для достижения приемлемой точности оказывается достаточным учесть 5 мнимых источников. Принимая во внимание, что итоговое распределение (3.1) есть линейная комбинация одномодальных распределений, для изучаемого распределения также можно ограничиться указанным числом мнимых источников для каждого из образующих его одномодальных распределений. Следовательно, размерность пространства оцениваемых параметров двумодального распределения оказывается равным 9.

3.2 Анализ точности оценивания параметров двумодальных распределений, получаемых в соответствии с алгоритмом, обоснованным в Главе 2

Для нахождения параметров двумодальных ФР и ПР СПООР с помощью ГА на первом этапе была изучена возможность применения алгоритма, описанному ранее в Главе 2.

Для этого, следуя разделу 2.1, было установлено взаимно однозначное соответствие между основными структурными элементами ГА, в том числе: видом элемента популяции (особи), оператор кроссовера (скрещивания), мутации, селекции (отбора), видом функции приспособленности (фитнес-функции) и параметрами изучаемого распределения $\mu_1, \sigma_1, a_1, b_1, \mu_2, \sigma_2, a_2, b_2, \alpha$ (рисунок 3.1).

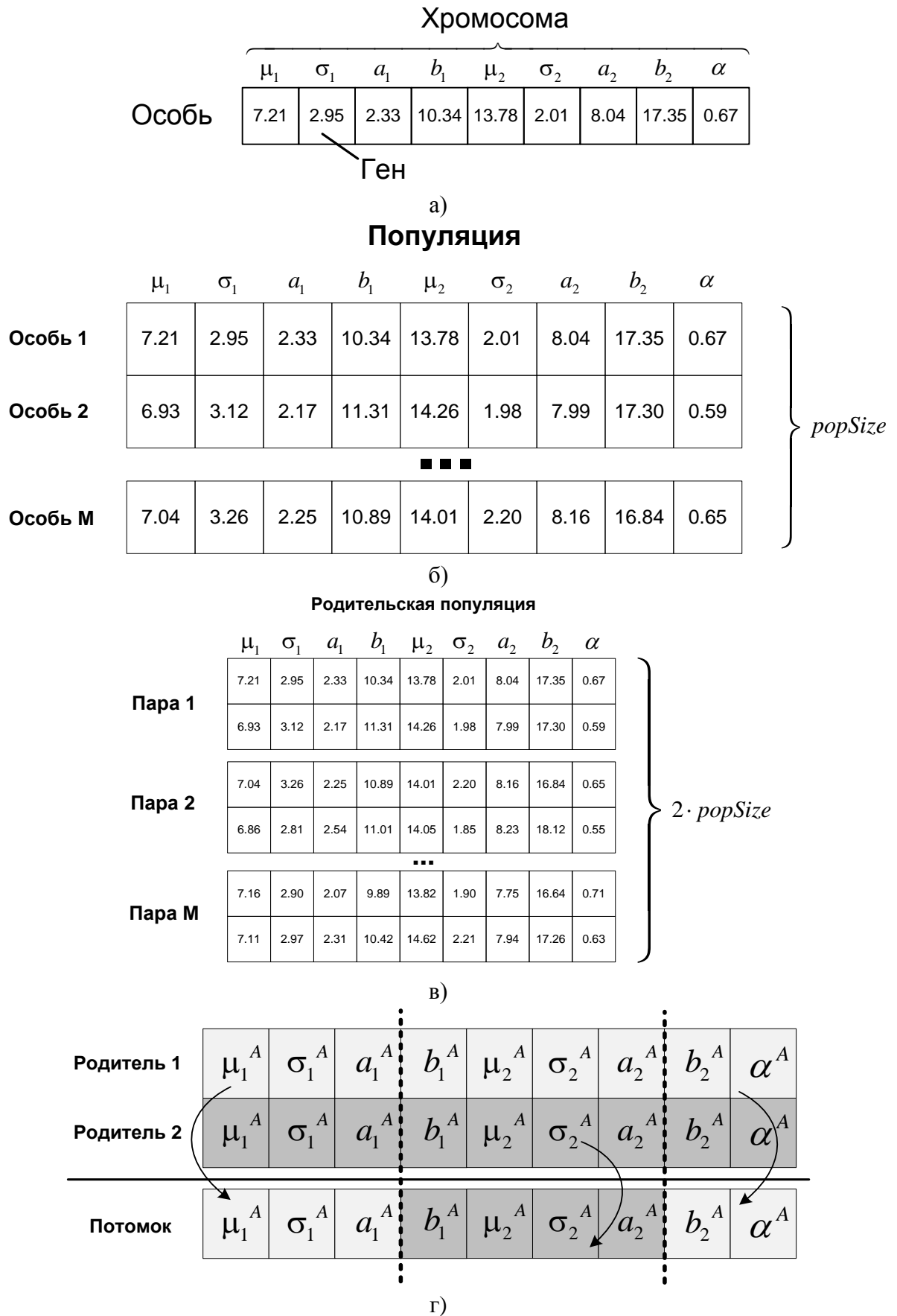


Рисунок 3.1. К установлению соответствия между параметрами распределения и параметрами ГА: а – представление особи, б – пример популяции, в – родительская популяция, полученная в результате селекции, г – пример двухточечного кроссовера

Результаты сравнения рисунка 3.1 и рисунка 2.2, позволяют сделать вывод о том, что для использования алгоритма оценивания ФР и ПР СПООР, описанного в предыдущем разделе, оказывается достаточным в использованном ранее алгоритме заменить хромосому $\{\mu, \sigma, x_{\min}, x_{\max}\}$ хромосомой $\{\mu_1, \sigma_1, a_1, b_1, \mu_2, \sigma_2, a_2, b_2, \alpha\}$.

При выборе границ областей поиска параметров двумодальных ФР и ПР СПООР мы руководствовались следующими правилами, основанными на результатах, полученных в разделе 2.2.

1. В качестве значений левой границ области рассеяния первой составляющей и правой границы второй составляющей случайной величины с двумодальным законом распределения в ограниченной области рассеяния X целесообразно использовать значения:

$$\begin{aligned}\tilde{a}_1 &= \min(X_i), \\ \tilde{b}_2 &= \max(X_i).\end{aligned}$$

2. Нулевое приближении параметров μ_1 и μ_2 следует выбирать из отрезка $[\tilde{a}_1, \tilde{b}_2]$, где \tilde{a}_1, \tilde{b}_2 – минимальное и максимальное значения случайной последовательности $X_i, i = 1, M$.

3. Область поиска истинных значений параметров μ_1 и μ_2 следует ограничить отрезком $[\tilde{a}_1, \tilde{b}_2]$.

4. Оценку нулевого приближения параметра σ следует получать на основе анализа зависимости

$$\langle x^2 \rangle = f(n), \quad (3.2)$$

где $\langle x^2 \rangle$ усредненное по ансамблю реализаций среднеквадратическое отклонение броуновской частицы, n – число шагов броуновской частицы, получаемой с помощью статистического моделирования броуновского движения в соответствие с алгоритмом, описанным в разделе 2.2.2.

5. Область поиска истинного значения параметра σ следует ограничить отрезком $[0, \sigma_{\max}]$, где $\sigma_{\max} = (\max f(n))^{0.5}$.

6. Область поиска истинного значения параметра α следует ограничить отрезком $[0, 1]$.

Анализ перечня действий, реализующих алгоритм решения рассматриваемой задачи с помощью ГА, показывает, что использованные подходы к оценке ряда значений параметров распределения (3.1) позволяют, фактически, уменьшить размерность пространства параметров с девяти до семи – $\mu_1, \sigma_1, b_1, \mu_2, \sigma_2, a_2, \alpha$.

В ходе проведенных исследований были изучены двумодальные распределения с ООР с параметрами, представленными в таблице 3.1.

Таблица 3.1. Параметры двумодальных распределений, использованных в исследовании

| Номер набора параметров | 1 | 2 | 3 | 4 | 5 |
|-------------------------|--------|--------|--------|--------|--------|
| μ_1 | 7.000 | 7.000 | 4.000 | 4.000 | 7.000 |
| σ_1 | 2.000 | 2.000 | 2.000 | 2.000 | 3.000 |
| a_1 | 3.000 | 3.000 | 3.000 | 3.000 | 3.000 |
| b_1 | 14.000 | 14.000 | 14.000 | 14.000 | 14.000 |
| μ_2 | 14.000 | 14.000 | 14.000 | 17.000 | 14.000 |
| σ_2 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| a_2 | 10.000 | 10.000 | 10.000 | 10.000 | 10.000 |
| b_2 | 18.000 | 18.000 | 18.000 | 18.000 | 18.000 |
| α | 0.500 | 0.700 | 0.600 | 0.500 | 0.700 |

Таким образом, было изучено $numSeq = 5$ различных распределений нормальных СПООР.

В исследовании использовались следующие настройки ГА:

- Селекция: $S = \langle \text{равномерная; турнирная; на основе рулетки} \rangle$;
- Мутация: $M = \langle \text{адаптивная} \rangle$ (т.к. в задаче присутствуют ограничения);
- Кроссовер: $C = \langle \text{одноточечный; двухточечный; усредненный; разбросанный, эвристический} \rangle$;
- Доля кроссовера: $Cfr = \langle 0,3; 0,6; 0,9 \rangle$;
- Доля мутации $1 - Cfr$;
- Размер популяции: $P = \langle 5, 10, 25 \rangle$.

Каждый эксперимент характеризовался следующим кортежем:

$$\langle S_s, M_m, C_c, Cfr_f, P_p \rangle,$$

поэтому всего было рассмотрено

$$numT = numS \times numM \times numC \times numCfr \times numP = 3 \times 1 \times 5 \times 3 \times 3 = 135$$

различных комбинаций настроек.

Для каждого набора настроек было выполнено $numExp = 10$ независимых статистических испытаний. Общее число запусков ГА составило

$$numTotal = numT \times numSeq \times numExp = 135 \times 5 \times 10 = 6750.$$

Наилучшие результаты работы ГА для каждого набора параметров распределения выбранных из результатов проведенных независимых испытаний представлены в таблице 3.2.

Таблица 3.2. Оценки параметров двумодального распределения с ограниченной областью рассеяния

| № набора параметров | Параметры | μ_1 | σ_1 | a_1 | b_1 | μ_2 | σ_2 | a_2 | b_2 | α | Δ_{integr} |
|---------------------|------------------------|---------|------------|-------|--------|---------|------------|--------|--------|----------|-----------------------|
| 1 | Теоретическое значения | 7.000 | 2.000 | 3.000 | 14.000 | 14.000 | 1.000 | 10.000 | 18.000 | 0.500 | $1.118 \cdot 10^{-1}$ |
| | Наилучший результат | 0.130 | 2.944 | 3.008 | 12.097 | 1.869 | 1.382 | 10.732 | 16.703 | 0.4710 | |
| 2 | Теоретическое значения | 7.000 | 2.000 | 3.000 | 14.000 | 14.000 | 1.000 | 10.000 | 18.000 | 0.700 | $9.403 \cdot 10^{-2}$ |
| | Наилучший результат | 0.056 | 1.673 | 3.112 | 13.595 | 0.960 | 4.059 | 7.888 | 15.983 | 0.545 | |
| 3 | Теоретическое значения | 4.000 | 2.000 | 3.000 | 14.000 | 14.000 | 1.000 | 10.000 | 18.000 | 0.600 | $1.992 \cdot 10^{-2}$ |
| | Наилучший результат | 1.502 | 2.083 | 3.008 | 11.011 | 0.756 | 1.170 | 7.337 | 17.512 | 0.608 | |
| 4 | Теоретическое значения | 4.000 | 2.000 | 3.000 | 14.000 | 17.000 | 1.000 | 10.000 | 18.000 | 0.500 | $6.801 \cdot 10^{-3}$ |
| | Наилучший результат | 2.597 | 2.262 | 3.008 | 14.237 | 1.526 | 1.457 | 9.633 | 17.962 | 0.524 | |
| 5 | Теоретическое значения | 7.000 | 3.000 | 3.000 | 14.000 | 14.000 | 1.000 | 10.000 | 18.000 | 0.700 | $1.112 \cdot 10^{-1}$ |
| | Наилучший результат | 0.518 | 2.785 | 3.023 | 11.714 | 2.733 | 2.785 | 7.868 | 16.493 | 0.514 | |

Гистограммы случайных последовательностей, теоретические функции плотности распределения и плотности распределения (3.1), вычисленные для значений параметров, приведенных в таблице 3.2, приведены на рисунке 3.2.

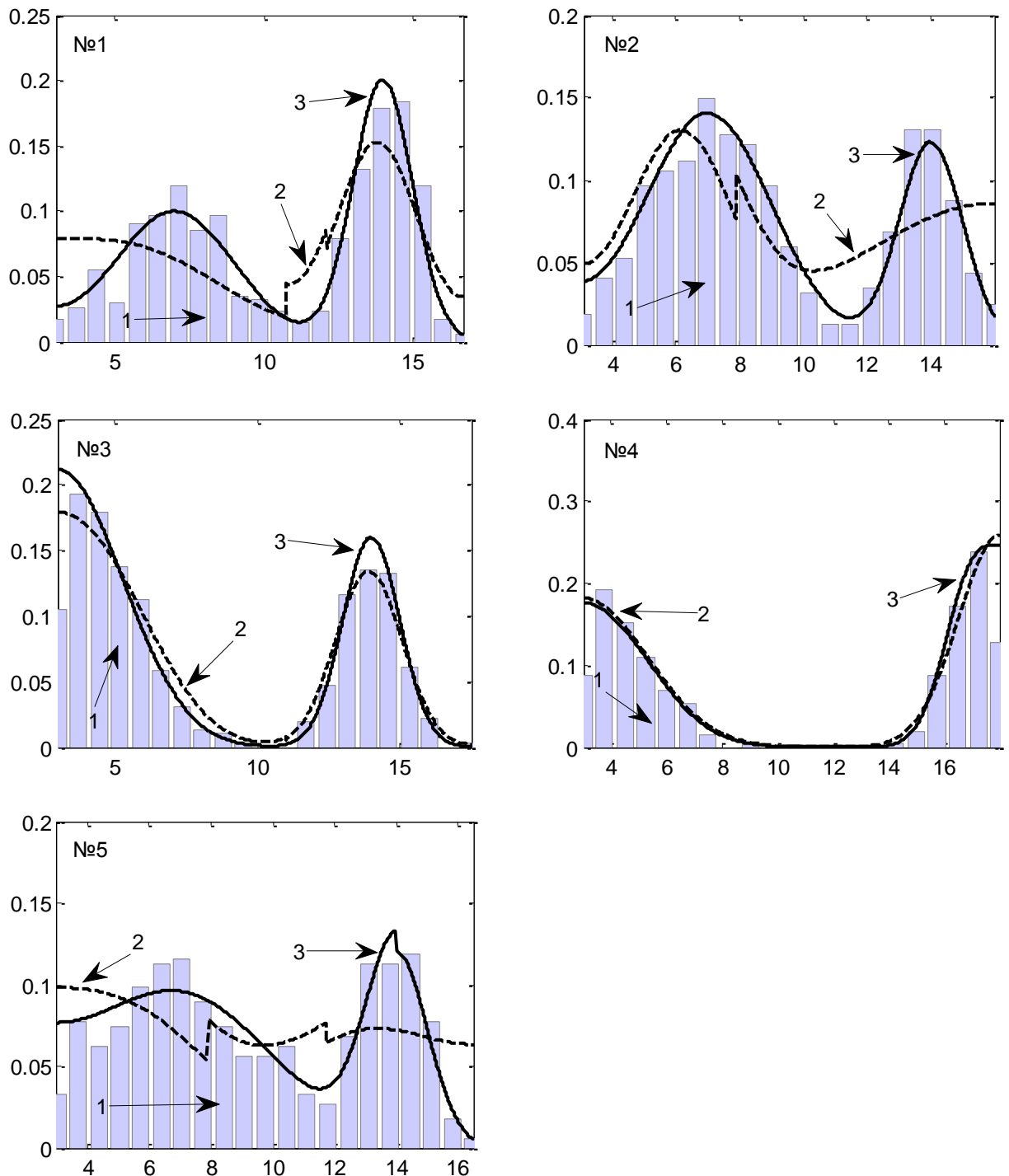


Рисунок 3.2. Функции плотности распределения, при вычислении которых использованы значения параметров, представленные в табл. 1 и настройки ГА, представленные в табл. 2:

1 – гистограмма случайной последовательности, 2 – экспериментальная функция плотности распределения, 3 – теоретическая функция плотности распределения

Из рисунка 3.2 и таблицы 3.2 видно, что точность аппроксимации функций плотности распределения случайной величины с двумодальным законом распределения в ограниченной области рассеяния для наборов

параметров 1, 2 и 5 оказывается недостаточно высокой. При этом интегральная погрешность на порядок больше, чем интегральная погрешность, полученная при анализе точности оценок параметров одномодальных распределений в разделе 2.3. Полученный результат, с нашей точки зрения, объясняется большой размерностью задачи (9 параметров), в сравнении с задачей, решавшейся в разделе 2.3. Следовательно, можно ожидать, что уменьшение количества параметров приведет к повышению точности нахождения параметров двумодального распределения с помощью ГА.

При этом понятно, что уменьшить число вычисляемых с помощью ГА параметров можно, если дополнительно привлечь методы, позволяющие получать независимые оценки одного или нескольких параметров двумодального распределения. В этой связи проведено исследование возможности получения данных параметров с помощью аппроксимации Розенблатта-Парзена, описание которого приведено в разделе 1.2.2, результаты которого обсуждаются в следующем разделе.

3.3 Исследование особенностей численных оценок параметра размытости в аппроксимации Розенблатта-Парзена

Напомним, следуя разделу 1.2.2, что для вычисления аппроксимации ФР и ПР случайной последовательности в методе Розенблатта-Парзена необходимо определить оптимальное значение параметра размытости h^* , являющееся решением уравнения (1.49). При этом найти искомого решение оказывается возможным только численно.

Отметим, что точность нахождения численного решения (1.49) напрямую зависит от удачного выбора начального приближения, общих правил для которого, однако, не существует. В [74] предложено на практике искать максимум функционала на основе анализа значений функции $\varphi(h_m)$, вычисляемых на интервале $[h_{m \min}, h_{m \max}]$. Однако при выполнении численных расчетов было обнаружено, что в ряде случаев для малых значений h функция $\varphi(h_m)$ оказывается негладкой и имеет разрывы первого рода [74], что затрудняет автоматизированный поиск абсциссы ее максимума. При этом объяснений причин обнаруженных особенностей функции $\varphi(h_m)$ не было дано, что потребовало предварить решение основной задачи целенаправленными исследованиями данной проблемы.

3.3.1 Анализ причин немонотонности информационного функционала у ядер с неограниченной областью определения

Рассмотрим результаты расчета значений в пакете MATLAB функции $\varphi(h)$ случайной последовательности $x_i, i = \overline{1, 100}$, члены которой принадлежат множеству действительных чисел, сгенерированной в соответствии с НР $N(1, 4)$, для ядра № 1 (рисунок 3.3).

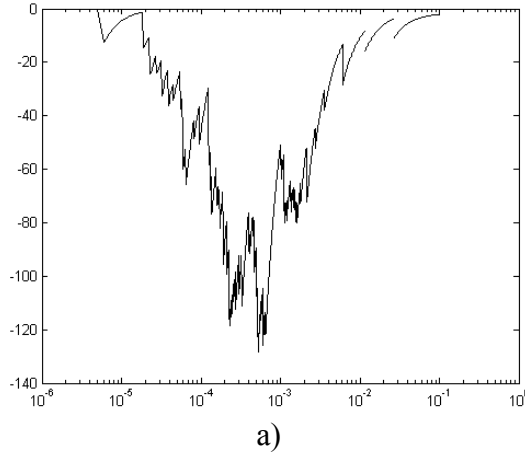


Рисунок 3.3. Зависимость $\varphi(h)$, ядро № 1, вычисленная для случайной последовательности $x_i, i = \overline{1, 100}$, сгенерированной в соответствии с нормальным законом распределения $N(1, 4)$

Из рисунка 3.3 видно, что при $h \in [10^{-4}, 10^{-1}]$ функция в точках $h \in [1.1643 \cdot 10^{-3}; 1.1645 \cdot 10^{-3}]$, $h \in [2.6196 \cdot 10^{-3}; 2.6230 \cdot 10^{-3}]$ имеет разрывы первого рода и второго рода. Для других значений параметра размытости h функция оказывается непрерывной. Анализ значений функции для значений параметра размытости h , находящихся в указанных диапазонах показал, что в точках разрывов второго рода, значения функции оказываются равными $-\infty$. Из (1.47) видно, что необходимым условием возникновения ситуации $\varphi(h) = -\infty$ является равенство нулю аргумента логарифма в (1.47):

$$\frac{1}{(Ns-1) \cdot h} \sum_{j \neq i}^{Ns-1} k \left(\frac{x_i - x_j}{h} \right) = 0.$$

Данный результат возможен при выполнении условия $k \left(\frac{x_i - x_j}{h} \right) \equiv 0$

одновременно для всех возможных сочетаний $i = \overline{1, 100}, j = \overline{1, 100}, i \neq j$. При аналитических вычислениях, как очевидно, данное условие для последовательности, состоящей из действительных чисел, не выполняется, так как не существует ни одной пары значений $i \neq j$, у которых $x_i = x_j$. При численных расчетах, напротив, для его выполнения достаточно, чтобы значения все вычисленные значения функции $k \left(\frac{x_i - x_j}{h} \right)$ оказались меньше машинного нуля $m_0 = 10^{-323}$. Отметим, что избежать выполнения данного условия и устранить разрывы второго рода оказывается возможным, проводя вычисления значений функции $\varphi(h)$ по следующей формуле:

$$\varphi(h) = -\ln(Ns-1) - \ln h + \sum_{i=1}^{Ns} \ln \left[\sum_{j \neq i}^{Ns-1} k_m \left(\frac{x_i - x_j}{h} \right) \right], \quad (3.3)$$

которая, как очевидно, эквивалентна (1.47). Результаты расчетов функции $\varphi(h)$ в соответствие с (3.3) представлены на рисунке 3.4, из которого видно, что, действительно, предложенный способ вычислений позволяет устранить разрывы второго рода.

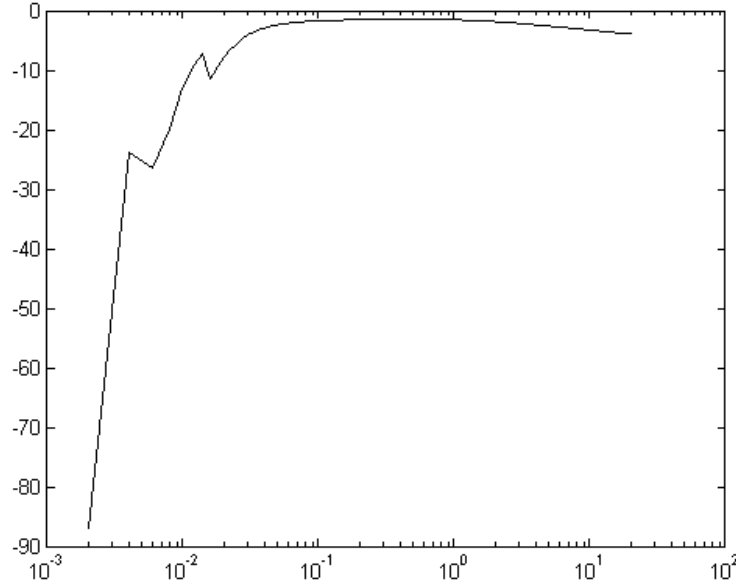


Рисунок 3.4. Зависимость $\varphi(h)$ для случайной последовательности $x_i, i = \overline{1, 100}$, сгенерированной в соответствие с нормальным законом распределения $N(1, 4)$, ядро № 1, вычисленная в соответствие с (3.3)

Приняв во внимание полученный результат, можно получить оценку минимально возможного значения параметра размытости для ядра № 1:

$$\exp\left(-\frac{(x_i - x_j)^2}{2h^2}\right) = 10^{-\lg e (x_i - x_j)^2 / 2h^2} > m_0,$$

откуда находим

$$h_{1\min} \geq \left(\frac{\min\left((x_i - x_j)^2 / 2\right)}{-\lg m_0 \cdot \ln 10} \right)^{1/2}.$$

Повторяя аналогичные рассуждения для ядер № 2, 4, получаем, соответственно:

$$h_{2\min} \geq \frac{\min(|x_i - x_j|)}{-\lg m_0 - \lg 2},$$

$$h_{4\min} \geq \frac{\min(x_i - x_j)^2}{1/(\pi \cdot m_0) - 1}.$$

Минимально возможное значение переменной $h_{5\min}$ можно найти как численное решение неравенства

$$2 \cdot \lg \left(1 + e^{-\min(|x_i - x_j|)/h} \right) + \frac{\min(|x_i - x_j|)}{h} \cdot \lg e \leq -\lg m_0.$$

Отметим, что можно изменить алгоритм поиска максимального значения $\varphi(h)$. Для этого заметим, что при

$$h_1 \geq \max \left((x_i - x_j)^2 / 2 \right)^{1/2}$$

для нормального ядра,

$$h_{2,4,5} \geq \max(|x_i - x_j|)$$

для ядер Лапласа, Коши и логистического ядра значения аргументов ядерной функции принадлежат интервалу $[0,1]$, области значений ядерных функций

принадлежат интервалам $\left[\frac{1}{\sqrt{2\pi}} e^{-1}, \frac{1}{\sqrt{2\pi}} \right], \left[\frac{1}{2} e^{-1}, \frac{1}{2} \right], \left[\frac{1}{2\pi}, \frac{1}{\pi} \right], \left[\frac{e^{-1}}{(1+e^{-1})^2}, \frac{1}{4} \right]$,

соответственно.

Таким образом, для указанных значений параметра h вычислительных проблем, связанных с точностью представления чисел в компьютере не возникает. Следовательно, можно, задавшись некоторым конечным шагом, начинать вычисления для конкретного ядра с соответствующего значения $h_{1,2,4,5}$ и далее двигаться в сторону меньших значений переменной h вплоть до достижения максимального значения функции $\varphi(h)$.

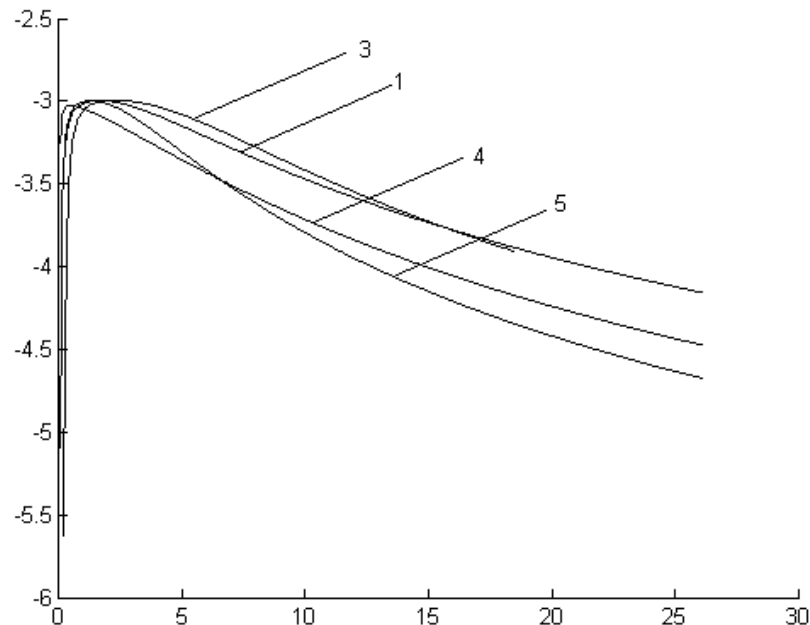


Рисунок 3.5. Зависимости $\varphi(h)$, ядра № 1, 2, 4, 5, вычисленные в соответствии с предложенным алгоритмом для случайной последовательности $x_i, i = \overline{1, 100}$, сгенерированной в соответствии с нормальным законом распределения $N(1, 4)$

Целесообразность использования на практике данного алгоритма подтверждается зависимостями, представленными на рисунке 3.5, которые, как очевидно, на выбранных интервалах оказываются гладкими.

3.3.2 Свойства информационного функционала для ядер с ограниченной областью определения

Ядра № 3, 6–9 имеют ограниченные области рассеяния: $\left|\frac{t}{2}\right| \leq \pi$, $|t| \leq \sqrt{5}$, $|t| \leq 1$, соответственно, поэтому

$$h_3 \geq \frac{\max(x_i - x_j)}{2\pi}, \quad h_6 \geq \frac{\max(x_i - x_j)}{\sqrt{5}}, \quad h_{7,8,9} \geq \max(x_i - x_j). \quad (3.4)$$

Результаты расчета значений функции $\varphi(h)$ для случайной последовательности $x_i, i = \overline{1, 100}$, сгенерированной в соответствии с нормальным законом распределения $N(1, 4)$, ядра № 3, 6–9, представлены на рисунке 3.6.

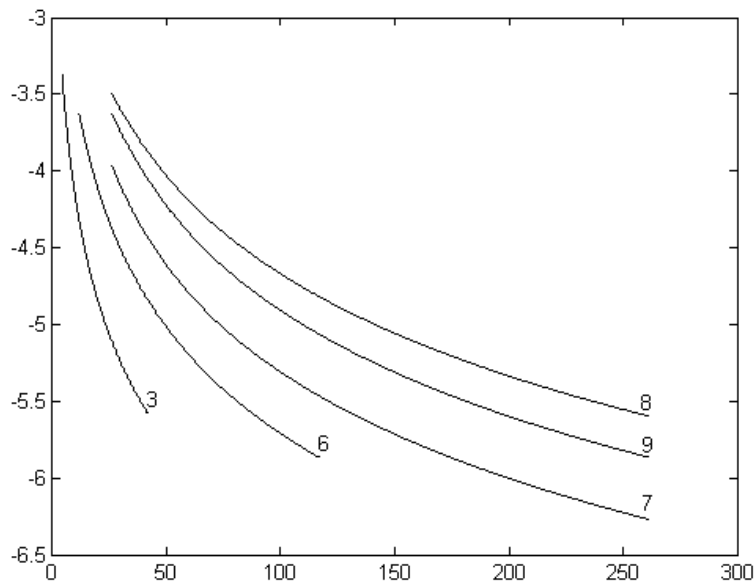


Рисунок 3.6. Зависимости $\varphi(h)$, ядра № 3, 6–9, вычисленные в соответствии с предложенным алгоритмом для случайной последовательности $x_i, i = \overline{1, 100}$, сгенерированной в соответствии с нормальным законом распределения $N(1, 4)$

Из рисунка 3.6 видно, что для каждого из использованных ядер изучаемая функция $\varphi(h)$ оказывается монотонно убывающей, поэтому для оценки оптимального значения параметра $h_{3,6-9}^*$ оказывается достаточным вычислить значение функции $\varphi(h)$ в точках, вычисляемых в соответствии с (3.4).

Таким образом, результаты проведенного исследования позволили:

1. объяснить причину, приводящую к немонотонной зависимости вычисленных значений функции $\varphi(h)$ от параметра h , обусловленной конечной точностью компьютерной арифметики;
2. для ядерных функций с неограниченной областью значений параметра h обосновать выбор начального значения и предложить алгоритм нахождения максимального значения функции $\varphi(h)$;
3. для ядерных функций с ограниченной областью значений параметра h обосновать выбор оптимального значения параметра h^* , обеспечивающего выполнение условия максимума информационного функционала (1.48).

3.4 Использование аппроксимации Розенблатта-Парзена для восстановления функции распределения дискретной случайной величины

Отметим, что формально метод Розенблатта-Парзена позволяет, построить аппроксимации ПР и ФР функции распределения любой конечной случайной последовательности, которая при условии правильного выбора параметра h , оказывается достаточно гладкой (см. [72]). При $Ns \rightarrow \infty, h_{Ns} \rightarrow 0$ оценка (1.44) является непараметрической и асимптотически оптимальной оценкой функции распределения в классе непрерывных функций [72]. Однако на практике случайные последовательности $x_i, i = \overline{1, Ns}$, для которых оцениваются их ФР и ПР, имеют конечную длину. В этой ситуации метод Розенблатта-Парзена, вообще говоря, не гарантирует близости аппроксимирующей функции, найденной с его помощью, к истинным ФР и ПР. В этой связи представляет практический интерес провести анализ точности восстановления методом Розенблатта-Парзена ФР случайных последовательностей конечной длины, теоретический закон распределения которых известен.

Для генерации случайных последовательностей мы использовали метод Монте-Карло, алгоритм реализации которого подробно описан в разделе 2.2.1.

3.4.1 Анализ проблемы, возникающей в методе Розенблатта-Парзена при оценке значения h^* дискретной последовательности, и способ ее устранения

Моделирование случайных блужданий броуновской частицы с ограниченной области рассеяния производилось в пакете MATLAB. При этом были использованы следующие параметры блуждания:

- число испытаний (количество элементов в последовательности) $Ns = 1000$;
- координаты начального положения броуновской частицы $x_0 = [0, 6, 12, 18]$;

- длина интервала рассеяния $2l = 48$;
- координаты границ интервала рассеяния $x_1 = -l, x_2 = l$;
- величина шага блужданий $step = 1$;
- количество шагов блужданий $Nsteps = [10, 50, 100, 500, 1000, 2000]$.

В ходе проведенных исследований было обнаружено, что в ряде случаев информационный функционал (1.48) случайной последовательности $x_i, i = \overline{1, 100}, x_0 = 0$ имеет два локальных экстремума (рисунок 3.7а), причем значение параметра размытости, обеспечивающее максимум информационного функционала не обеспечивает оптимальное качество аппроксимации (рисунок 3.7б).

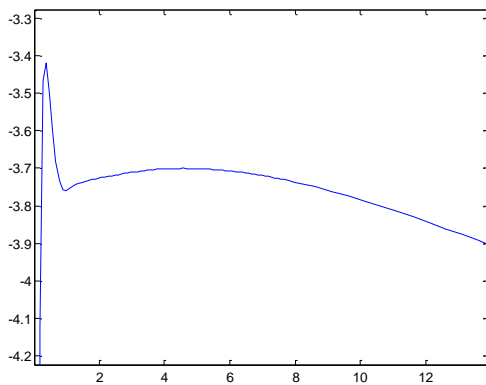


Рисунок 3.7а. Зависимость информационного функционала от параметра h (ядро № 1, 1.7)

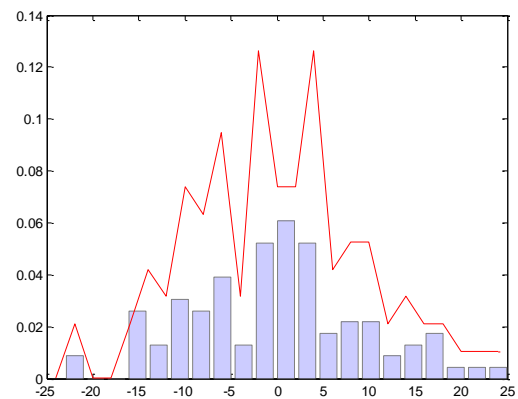


Рисунок 3.7б. Аппроксимация Розенблатта-Парзена, значение параметра размытости $h = 0.4$

Анализ случайной последовательности $x_i, i = \overline{1, 100}, x_0 = 0$ показал, что члены данной последовательности представляют собой набор целочисленных значений, повторяющихся с ненулевой вероятностью, поэтому для некоторых сочетаний $i \neq j$ условие $x_i \neq x_j$ оказывается невыполненным. Это отличает данную последовательность от последовательностей с континуальным (квазинепрерывным) множеством значений, у которых выполнение условия $i \neq j$ эквивалентно выполнению условия $x_i \neq x_j$. В этой связи в формуле (1.48) следует заменить условие $i \neq j$ условием $x_i \neq x_j$:

$$\varphi(h_m) = \frac{1}{N_s} \sum_{i=1}^{N_s} \ln \left[\frac{1}{(N_s - 1) \cdot h_m} \sum_{\substack{j=1 \\ x_i \neq x_j}}^{N_s-1} k_m \left(\frac{x_i - x_j}{h_m} \right) \right]. \quad (3.5)$$

Результаты расчета значений информационного функционала в соответствии с (3.5) и аппроксимации Розенблатта-Парзена представлены на рисунке 3.8, из которого видно, что предложенная модификация расчетной формулы, действительно, устраняет проблему появления ложного максимума и, следовательно, обеспечивает нахождение истинного значения h^* .

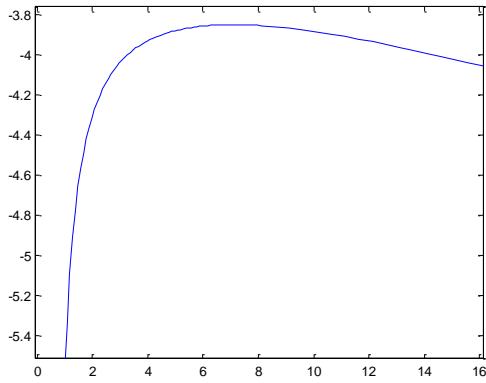


Рисунок 3.8а. Зависимость информационного функционала (3.5) от параметра h (ядро № 1, 1.7)

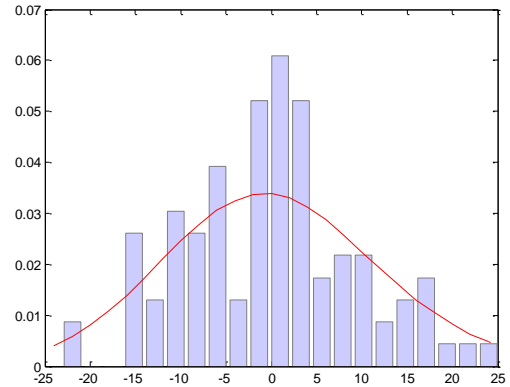
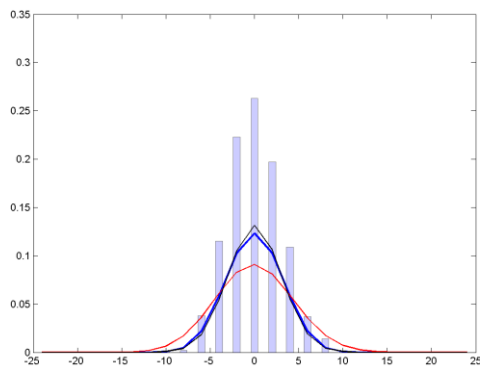


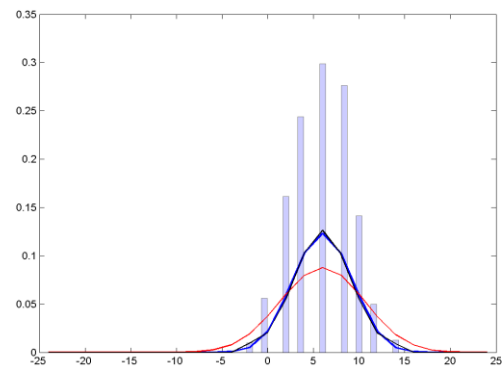
Рисунок 3.8б. Аппроксимация Розенблатта-Парзена, значение параметра размытости $h = 7.0$.

3.4.2 Сравнение качества аппроксимации функции распределения дискретной случайной последовательности методом Розенблатта-Парзена и методом мнимых источников

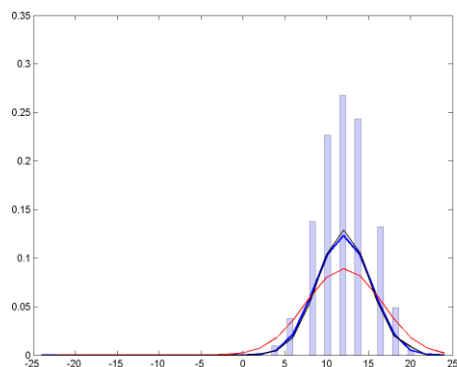
Типичные результаты (гистограммы, рассчитанные по аналитическим формулам, случайных последовательностей, сгенерированных в соответствие с методикой, описанной в предыдущем разделе; графики функций $F_{RP}(x)$, $F_{ISM}(x)$ и $F_{Teor}(x)$) представлены на рисунке 3.9.



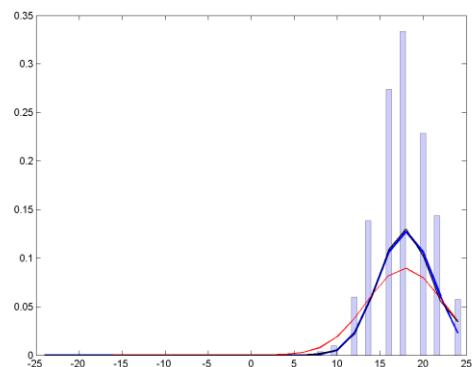
а)



б)



в)



г)

Рисунок 3.9. Результаты аппроксимации ФР броуновской частицы в ограниченной области рассеяния: а) $x_0 = 0, n = 10$; б) $x_0 = 6, n = 10$; в) $x_0 = 12, n = 10$; г) $x_0 = 18, n = 10$

В качестве количественной меры отличия эмпирических ФР $F_{RP}(x)$, $F_{ISM}(x)$ от теоретической ФР $F_{Teor}(x)$ были использованы среднеквадратические отклонения остатков каждой из построенных моделей плотности распределения:

$$\Delta_{RP} = \frac{1}{Ns-1} \sum_{i=1}^{Ns} (F_{Teor}(x_i) - F_{RP}(x_i))^2,$$

$$\Delta_{RP} = \frac{1}{Ns-1} \sum_{i=1}^{Ns} (F_{Teor}(x_i) - F_{ISM}(x_i))^2,$$

представленные в таблицах 3.3, 3.4.

Таблица 3.3. Погрешность метода мнимых источников

| Метод мнимых источников | | Количество шагов блужданий | | | | | |
|-------------------------------|----|----------------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| | | 10 | 50 | 100 | 500 | 1000 | 2000 |
| Точка отсчета | 0 | $5,551 \cdot 10^{-6}$ | $1,929 \cdot 10^{-6}$ | $1,855 \cdot 10^{-6}$ | $9,571 \cdot 10^{-6}$ | $2,538 \cdot 10^{-6}$ | $8,272 \cdot 10^{-9}$ |
| | 6 | $3,151 \cdot 10^{-6}$ | $2,957 \cdot 10^{-7}$ | $1,935 \cdot 10^{-6}$ | $3,195 \cdot 10^{-6}$ | $9,215 \cdot 10^{-7}$ | $1,089 \cdot 10^{-6}$ |
| | 12 | $3,912 \cdot 10^{-6}$ | $7,761 \cdot 10^{-6}$ | $2,375 \cdot 10^{-5}$ | $6,101 \cdot 10^{-6}$ | $1,839 \cdot 10^{-6}$ | $8,535 \cdot 10^{-8}$ |
| | 18 | $7,997 \cdot 10^{-6}$ | $4,641 \cdot 10^{-5}$ | $3,441 \cdot 10^{-5}$ | $3,481 \cdot 10^{-6}$ | $1,695 \cdot 10^{-6}$ | $3,223 \cdot 10^{-6}$ |

Таблица 3.4. Погрешность метода Розенблатта-Парзена

| Метод Розенблатта- Парзена | | Количество шагов блужданий | | | | | |
|----------------------------------|----|----------------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| | | 10 | 50 | 100 | 500 | 1000 | 2000 |
| Точка отсчета | 0 | $1,067 \cdot 10^{-5}$ | $3,028 \cdot 10^{-5}$ | $1,360 \cdot 10^{-5}$ | $6,766 \cdot 10^{-6}$ | $1,085 \cdot 10^{-5}$ | $1,676 \cdot 10^{-5}$ |
| | 6 | $1,295 \cdot 10^{-4}$ | $2,773 \cdot 10^{-5}$ | $1,236 \cdot 10^{-5}$ | $6,656 \cdot 10^{-6}$ | $1,185 \cdot 10^{-5}$ | $1,831 \cdot 10^{-5}$ |
| | 12 | $1,191 \cdot 10^{-4}$ | $2,837 \cdot 10^{-5}$ | $1,337 \cdot 10^{-5}$ | $9,189 \cdot 10^{-6}$ | $1,537 \cdot 10^{-5}$ | $1,660 \cdot 10^{-5}$ |
| | 18 | $1,328 \cdot 10^{-4}$ | $3,764 \cdot 10^{-5}$ | $2,247 \cdot 10^{-5}$ | $1,627 \cdot 10^{-6}$ | $1,812 \cdot 10^{-5}$ | $1,759 \cdot 10^{-5}$ |

Из таблиц 3.3 и 3.4 видно, что аппроксимация ФР случайных броуновских блужданий в ограниченной области рассеяния методом мнимых источников в подавляющем большинстве случаев оказывается более точной, чем аппроксимация Розенблатта-Парзена. Данный результат, с нашей точки зрения, объясняется тем, что, метод мнимых источников в отличие от аппроксимации Розенблатта-Парзена основан на физической модели изучаемого процесса, а потому оказывается более адаптированным к особенностям изучаемого распределения.

3.4.3 Выводы

1. В ходе изучения возможности применения аппроксимации Розенблатта-Парзена в задаче аппроксимации ФР последовательностей, представляющих собой значения координат броуновской частицы, совершающей случайные блуждания в ограниченной области рассеяния, обнаружено, что зависимость информационного функционала (1.48) от параметра размытости, вычисляемая в соответствие со стандартным

алгоритмом метода Розенблатта-Парзена, в ряде случаев может иметь два локальных максимума, что затрудняет нахождение истинного оптимального значения параметра размытости h^* .

2. Предложена модификация алгоритма нахождения оптимального значения параметра размытости h^* , пригодная для дискретных целочисленных случайных последовательностей.

3. Проведено сравнение точности аппроксимации ФР случайной величины с ограниченной областью, найденное методом Розенблатта-Парзена, и методом мнимых источников, с теоретическим законом распределения, результаты которого показывают, что у метода мнимых источников, основанного на физической модели изучаемого процесса, погрешность оказывается меньше, а потому в рассмотренной задаче целесообразно использовать метод мнимых источников.

3.5 Использование аппроксимации Розенблатта-Парзена для восстановления функции распределения непрерывной случайной величины с ограниченным одномодальным законом распределения

Проведем анализ точности восстановления ФР непрерывной одномодальной случайной величины с ограниченной областью рассеяния методом Розенблатта-Парзена и методом мнимых источников, в котором параметры аппроксимирующей функции находятся с помощью ГА.

3.5.1 Методика исследования

Для анализа точности восстановления ФР одномодальной случайной величины с ограниченной областью рассеяния выбранными методами мы использовали методику, реализующуюся следующей последовательностью действий.

1) В соответствии с (2.1) генерировались методом обратного преобразования случайные последовательности $\{x_i\}$, $i = \overline{1, N_s}$.

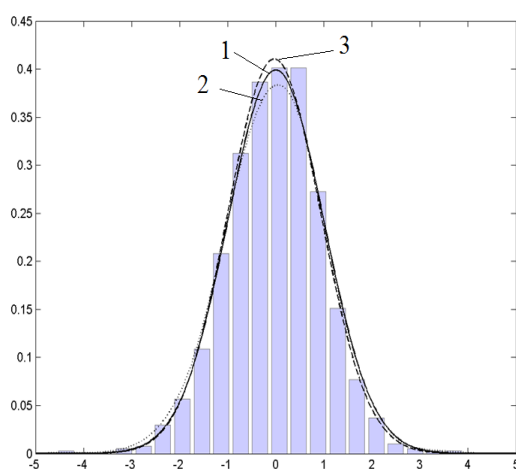
2) Далее для каждой последовательности $\{x_i\}$ находились эмпирические оценки функций распределения и плотностей распределения методом Розенблатта-Парзена ($F_{RP}(x)$ и $f_{RP}(x)$, соответственно) и методом мнимых источников ($F_{IMS}(x)$ и $f_{RP}(x)$, соответственно), которые сравнивались далее с известными теоретическими функцией распределения $F_{Teor}(x)$ и плотностью распределения $f_{teor}(x)$.

3.5.2 Анализ точности аппроксимации функции распределения случайной последовательности методом Розенблатта-Парзена

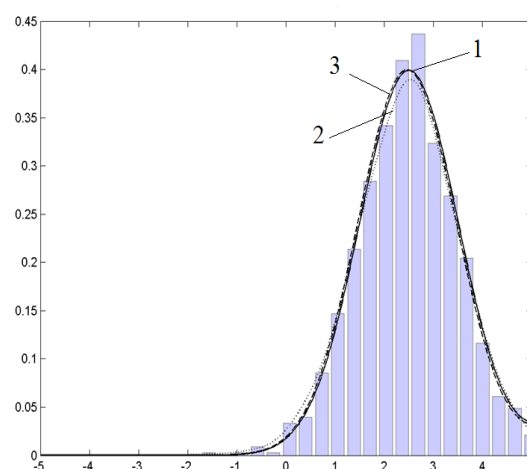
Рассмотрим типичные результаты, полученные для следующих параметров блуждания:

- количество элементов в случайной последовательности $N_s = 1000$;
- размер области рассеяния $[-5; 5]$
- координаты центра распределения $\mu = [-5, -2.5, 0, 2.5, 5]$;
- СКО случайной величины $\sigma = [1, 3, 5, 7, 10]$,

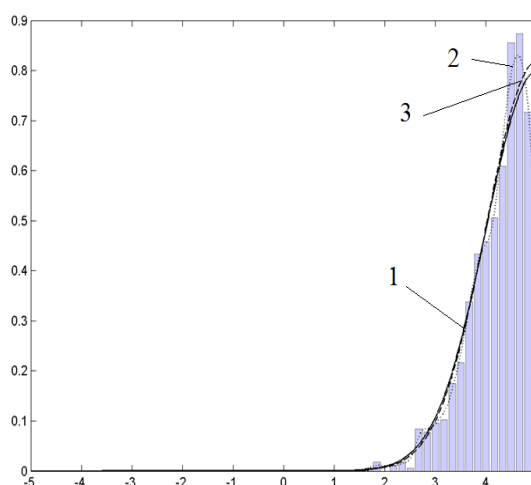
представленные на рисунках 3.10, 3.11.



а)

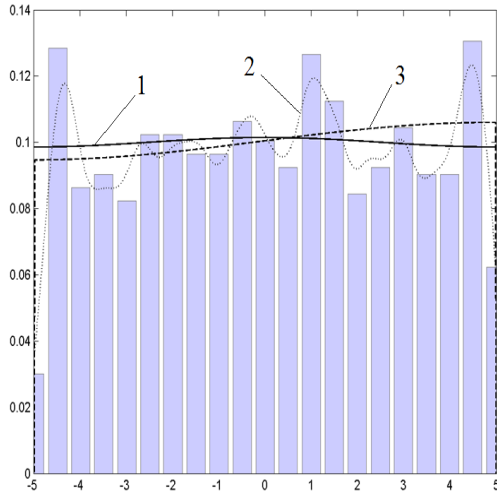


б)

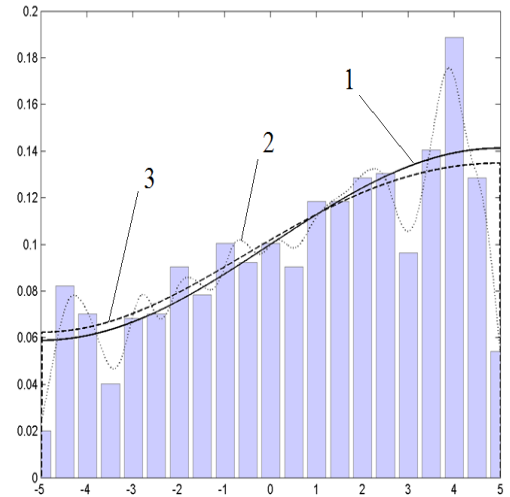


в)

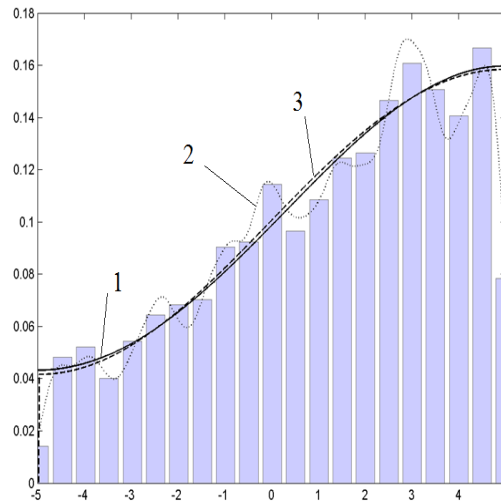
Рисунок 3.10. Результаты аппроксимации ПР случайной последовательности $\{x_i\}$: а) $\mu=0, \sigma=1$; б) $\mu=2.5; \sigma=1$ в) $\mu=5, \sigma=1$; 1 – $f_{\text{теор}}(x)$; 2 – $f_{RP}(x)$, ядерная функция №5, №2, №1 (см. 1.7), соответственно; 3 – $f_{IMS}(x)$



а)



б)



в)

Рисунок 3.11. Результаты аппроксимации ПР случайной последовательности $\{x_i\}$: а) $\mu = 0, \sigma = 5$; б) $\mu = 2.5, \sigma = 5$; в) $\mu = 5, \sigma = 5$; 1 – $f_{teor}(x)$; 2 – $f_{RP}(x)$, ядерная функция №3, №1, №3 (см. 1.7), соответственно; 3 – $f_{IMS}(x)$

В качестве количественной меры отличия эмпирических функций распределения $F_{RP}(x)$, $F_{IMS}(x)$ от теоретической функции распределения $F_{Teor}(x)$ были использовано среднеквадратические отклонения остатков моделей от теоретической плотности распределения:

$$\Delta_{RP} = \frac{1}{N_S - 1} \sum_{i=1}^{N_S} (F_{Teor}(x_i) - F_{RP}(x_i))^2,$$

$$\Delta_{IMS} = \frac{1}{N_S - 1} \sum_{i=1}^{N_S} (F_{Teor}(x_i) - F_{IMS}(x_i))^2,$$

представленные в таблицах 3.5, 3.6.

Таблица 3.5. Погрешность оценки функции распределения методом Розенблатта-Парзена

| Метод Розенблатта-Парзена | | Центр распределения, μ | | | | |
|---------------------------|----|----------------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| | | -5 | -2,5 | 0 | 2,5 | 5 |
| СКО, σ | 1 | $1,432 \cdot 10^{-3}$ | $3,531 \cdot 10^{-5}$ | $1,389 \cdot 10^{-4}$ | $7,225 \cdot 10^{-5}$ | $1,428 \cdot 10^{-3}$ |
| | 3 | $2,274 \cdot 10^{-4}$ | $2,608 \cdot 10^{-4}$ | $8,821 \cdot 10^{-5}$ | $1,752 \cdot 10^{-4}$ | $5,532 \cdot 10^{-4}$ |
| | 5 | $2,792 \cdot 10^{-4}$ | $1,855 \cdot 10^{-4}$ | $1,638 \cdot 10^{-4}$ | $1,616 \cdot 10^{-4}$ | $2,011 \cdot 10^{-4}$ |
| | 7 | $2,004 \cdot 10^{-4}$ | $1,440 \cdot 10^{-4}$ | $1,298 \cdot 10^{-4}$ | $1,093 \cdot 10^{-4}$ | $1,518 \cdot 10^{-4}$ |
| | 10 | $1,431 \cdot 10^{-4}$ | $1,316 \cdot 10^{-4}$ | $1,379 \cdot 10^{-4}$ | $1,727 \cdot 10^{-4}$ | $1,470 \cdot 10^{-4}$ |

Таблица 3.6. Погрешность оценки функции распределения методом мнимых источников

| Метод мнимых источников | | Центр распределения, μ | | | | |
|-------------------------|----|----------------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| | | -5 | -2,5 | 0 | 2,5 | 5 |
| СКО, σ | 1 | $6,412 \cdot 10^{-4}$ | $3,423 \cdot 10^{-5}$ | $1,498 \cdot 10^{-5}$ | $7,267 \cdot 10^{-5}$ | $1,125 \cdot 10^{-4}$ |
| | 3 | $7,934 \cdot 10^{-5}$ | $2,924 \cdot 10^{-5}$ | $2,274 \cdot 10^{-5}$ | $2,917 \cdot 10^{-5}$ | $9,254 \cdot 10^{-5}$ |
| | 5 | $2,315 \cdot 10^{-5}$ | $3,184 \cdot 10^{-5}$ | $4,368 \cdot 10^{-5}$ | $6,254 \cdot 10^{-5}$ | $4,132 \cdot 10^{-5}$ |
| | 7 | $3,157 \cdot 10^{-5}$ | $8,193 \cdot 10^{-5}$ | $2,141 \cdot 10^{-5}$ | $5,432 \cdot 10^{-5}$ | $6,778 \cdot 10^{-5}$ |
| | 10 | $5,682 \cdot 10^{-5}$ | $6,647 \cdot 10^{-5}$ | $2,798 \cdot 10^{-5}$ | $1,782 \cdot 10^{-5}$ | $1,167 \cdot 10^{-5}$ |

Из таблиц 3.5 и 3.6 видно, что аппроксимация ФР непрерывной случайной величины в ограниченной области рассеяния методом мнимых источников для большинства рассмотренных значений параметров оказывается более точной, чем аппроксимация Розенблатта-Парзена. Данный результат, с нашей точки зрения, объясняется следующим. Аппроксимация Розенблатта-Парзена при правильном выборе значения параметра размытости h гарантирует получение гладких оценок функции распределения. Однако данный метод является, в известной мере, формальным, поскольку выбор ядерной функции оказывается независимым от математической модели изучаемого процесса. В то время как в методе мнимых источников математическая модель распределения является решением уравнения случайных блужданий броуновской частицы в ограниченной области рассеяния, а потому более адаптированным к особенностям распределения изучаемых случайных последовательностей $\{x_i\}$.

Также проведено сравнение времен, затрачиваемых для вычисления аппроксимаций ФР случайных последовательностей каждым из методов. Оказалось, что в программной реализации метода мнимых источников с применением ГА, использующегося для поиска параметров распределения, одна итерация выполняется ≈ 5 секунд. При этом параметры распределения, обеспечивающие минимальное значение целевой функции генетического алгоритма, находится за 4 итерации. Таким образом, общее время

нахождения решения методом мнимых источников составляет ≈ 20 с. В методе Розенблатта-Парзена наиболее затратной по времени оказывается операция нахождения максимума информационного функционала (1.48) для ядер с неограниченной областью определения № 1, 2, 4, 5. Здесь для нахождения значения h_m^* в соответствии с (1.47) приходится вычислять значения функции $\varphi(h_m)$ в соответствии с (1.48) для нескольких значений h_m , т.е. для каждого значения h_m , приходится вычислять в (1.48) двойную сумму по индексам i и j , поэтому в зависимости от типа ядра затрачивается от 140 до 200 секунд при условии удачного выбора интервала поиска h_m^* (см. подробнее раздел 3.4).

Так как для ограниченных ядер получены простые с вычислительной точки зрения аналитические выражения для вычисления оптимального значения параметра размытости h_m^* , поэтому время вычисления является незначительным. В результате, общее время нахождения аппроксимации функции распределения случайной последовательности $\{x_i\}$ методом Розенблатта-Парзена составляет $\approx 640-800$ секунд. Таким образом, в рассматриваемой задаче временные затраты при использовании метода мнимых источников оказывают в $\approx 32-40$ раза меньше, чем у метода Розенблатта-Парзена.

3.5.3 Выводы

Проведено сравнение точности аппроксимации функции распределения непрерывной случайной величины с ограниченной областью, вычисляемого методом Розенблатта-Парзена и методом мнимых источников, с теоретическим законом распределения. Результаты сравнения показывают, что в рассматриваемой задаче метод мнимых источников, основанный на физической модели изучаемого процесса, оказывается предпочтительнее как с точностной, так и вычислительной точек зрения.

3.6 Методика оценивания параметров случайной величины со смешанным двумодальным законом распределения на основе совместного использования аппроксимации Розенблатта-Парзена, метода мнимых источников и ГА

Для оценки возможности уменьшения размерности задачи для ГА с помощью использования метода Розенблатта-Парзена, позволяющего аппроксимировать функции плотности распределения [121,117], рассмотрим типичные результаты аппроксимации Розенблатта-Парзена двумодальной ПР СПООР, представленные на рисунке 3.12.

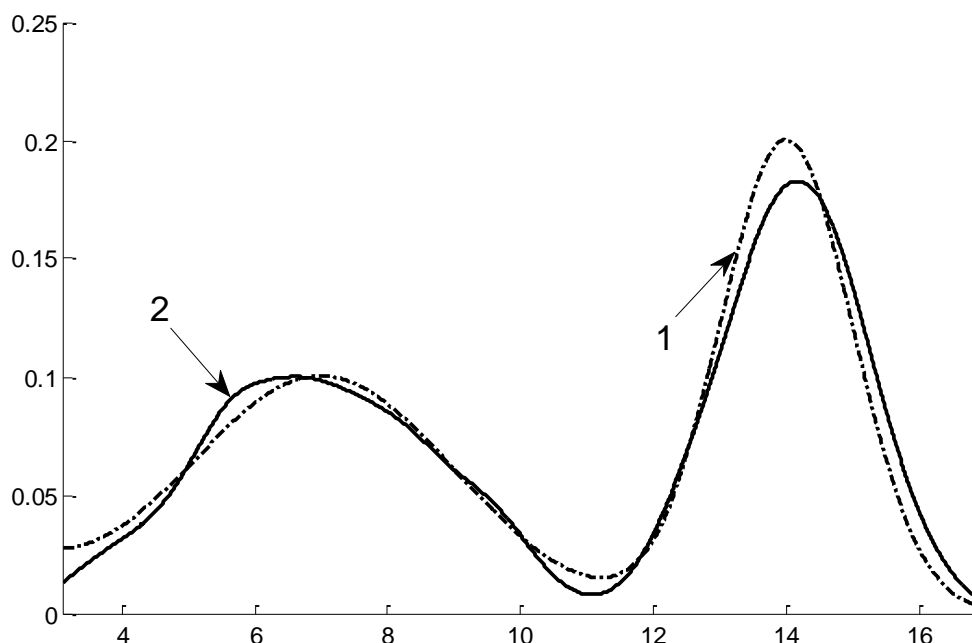


Рисунок 3.12. Двумодальная ПР: 1 – теоретическая ПР; 2 – аппроксимация Розенблатта-Парзена ПР

Из рисунка 3.12 видно, что на основе анализа аппроксимирующей кривой можно получать оценки математических ожиданий каждой из мод исходного распределения $\mu_{1,2}$, а также значение левой границы первой составляющей a_1 и правой границы второй составляющей b_2 . Если использовать эти значения в ГА для поиска оставшихся параметров двумодального распределения, то с помощью ГА остается вычислить следующие параметры: $\sigma_1, b_1, \sigma_2, a_2, \alpha$, следовательно, удастся уменьшить размерность задачи с 9 до 5. Далее, зная значения параметров $\sigma_1, b_1, \sigma_2, a_2, \alpha$, можно уточнить с помощью ГА значения параметров $\mu_{1,2}, a_1, b_2$ и затем повторить описанную выше процедуру. При этом можно ожидать, что точность значений оцениваемых параметров будет повышаться.

Таким образом, методика нахождения параметров двумодальной ПР СПООР $\{x_i\}, i = \overline{1, Ns}$ реализуется выполнением следующей последовательности действий:

1. Вычисление в соответствии с методом Розенблатта-Парзена значений функции, аппроксимирующей ПР (3.1), – $f_{RP}(x)$.

2. Вычисление оценок значений параметров $\tilde{\mu}_1, \tilde{\mu}_2$ – абсцисс локальных максимумов функции $f_{RP}(x)$.

3. Вычисление оценки координаты левой границы области рассеяния моды распределения (3.1), описываемой функцией $F_1(x, \mu_1, \sigma_1, a_1, b_1)$: $\tilde{a}_1 = \min_i (X_i)$.

4. Вычисление оценки координаты правой границы области рассеяния моды распределения (3.1), описываемой функцией $F_2(x, \mu_2, \sigma_2, a_2, b_2)$: $\tilde{b}_2 = \max_i (X_i)$.

5. Вычисление в соответствии с ГА значений параметров $\sigma_1, b_1, \sigma_2, a_2, \alpha$.

В ходе проведенных исследований описанного выше алгоритма мы использовали следующие настройки ГА:

- селекция: $S = \langle \text{равномерная; турнирная; на основе рулетки} \rangle$;
- мутация: $M = \langle \text{адаптивная} \rangle$, т.к. в задаче присутствуют ограничения;
- кроссовер: $C = \langle \text{одноточечный; двухточечный; усредненный; разбросанный, эвристический} \rangle$;
- доля кроссовера: $Cfr = \langle 0,3; 0,6; 0,9 \rangle$ (оставшаяся часть приходилась на мутацию);
- размер популяции: $P = \langle 5, 15, 30 \rangle$.

Таким образом, каждому из проведенных экспериментов можно поставить в соответствие определенный набор настроек – кортеж $\langle S_s, M_m, C_c, Cfr_f, P_p \rangle$. Всего в проведенных экспериментах было рассмотрено

$$numT = numS \times numM \times numC \times numCfr \times numP = 3 \times 1 \times 5 \times 3 \times 3 = 135$$

различных комбинаций настроек. Для каждого набора настроек было выполнено 50 независимых испытаний. Для повышения достоверности в качестве значений параметров $\tilde{\sigma}_1, \tilde{b}_1, \tilde{\sigma}_2, \tilde{a}_2, \tilde{\alpha}$ принимались средние по ансамблю испытаний значения.

Для количественной оценки качества найденных значений параметров $\tilde{\sigma}_1, \tilde{b}_1, \tilde{\sigma}_2, \tilde{a}_2$ решения мы использовали величину Δ_x :

$$\Delta_y = \frac{\max |y_{teor} - \tilde{y}|}{L} \cdot 100\%, \quad (3.6)$$

где $\tilde{y} = \tilde{\sigma}_1, \tilde{b}_1, \tilde{\sigma}_2, \tilde{a}_2$

Для количественной оценки качества найденного значения параметра $\tilde{\alpha}$ мы использовали величину Δ_α :

$$\Delta_\alpha = \frac{|\alpha_{teor} - \tilde{\alpha}|}{\alpha_{\max}} \cdot 100\%. \quad (3.7)$$

где $\alpha_{\max} = 1$.

Результаты расчетов погрешностей параметров $\tilde{\sigma}_1, \tilde{b}_1, \tilde{\sigma}_2, \tilde{a}_2, \tilde{\alpha}$ в соответствие с (3.6), (3.7) для тестовых последовательностей, генерируемых в соответствие с распределением (3.1), где $\mu_1 = 7, \sigma_1 = 2, a_1 = 3, b_1 = 14, \mu_2 = 14, \sigma_2 = 1, a_2 = 6, b_2 = 18, \alpha = 0.5$ представлены на рисунке 3.13.

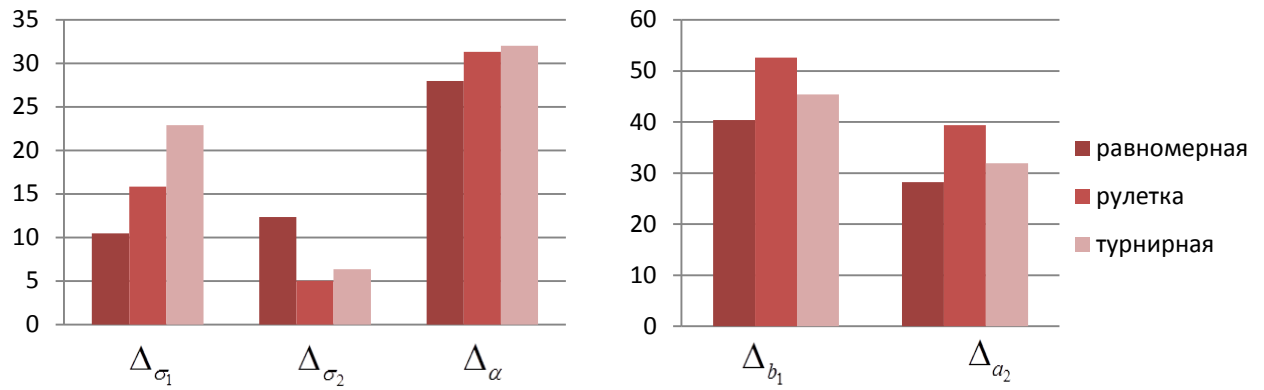


Рисунок 3.13а. Погрешности в разрезе вида селекции – равномерной, на основе рулетки и турнирной селекции

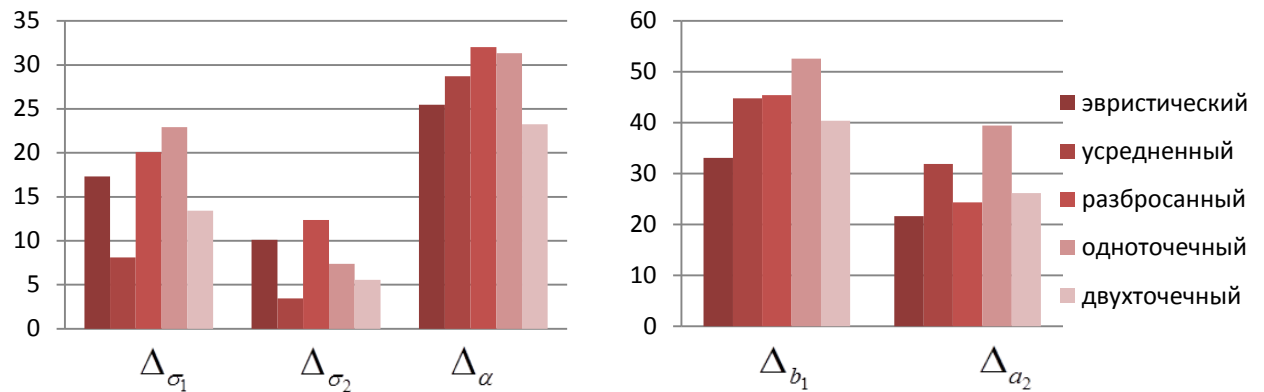


Рисунок 3.13б. Погрешности в разрезе вида кроссовера – эвристический, усредненный, разбросанный, одноточечный и двухточечный

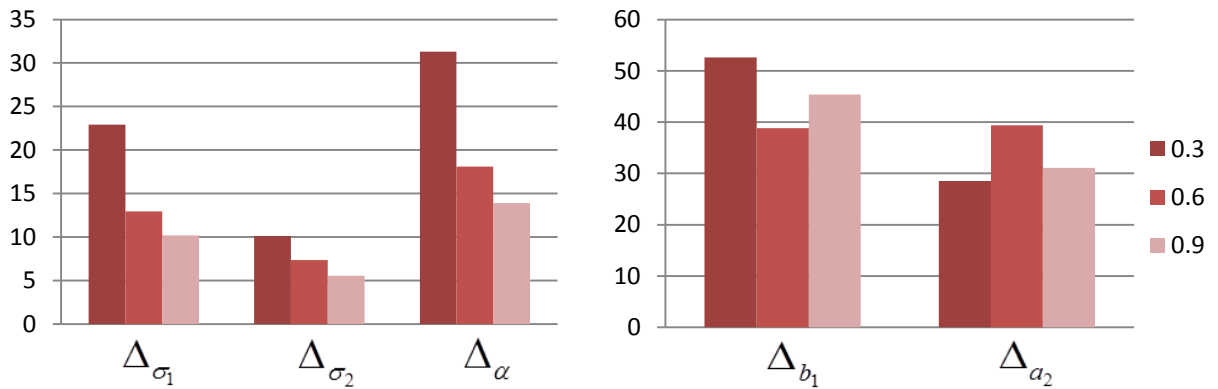


Рисунок 3.13в. Погрешности в разрезе доли кроссовера – 0.3, 0.6 и 0.9

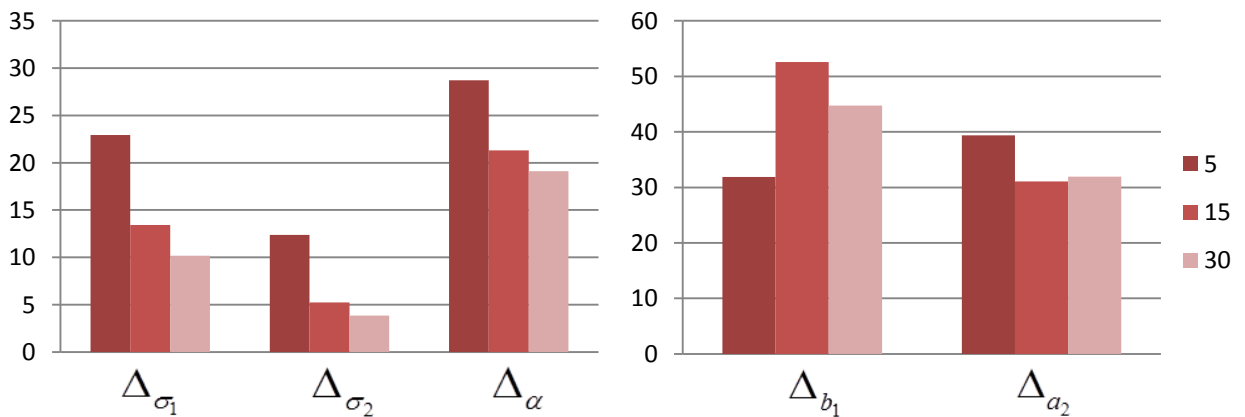


Рисунок 3.13г. Погрешности в разрезе размера популяции – 5, 15 и 30 особей

Из рисунка 3.13 видно, что виды селекции и кроссовера не оказывают существенного влияния на погрешность вычисления параметров $\tilde{\sigma}_1, \tilde{b}_1, \tilde{\sigma}_2, \tilde{a}_2, \tilde{\alpha}$. В то время как при увеличении доли кроссовера и числа особей в популяции точность вычисления ключевых параметров $\sigma_1, \sigma_2, \alpha$ увеличивается при одновременном увеличении времени расчетов. Приведем средние значения погрешностей параметров

$$\overline{\Delta_{\sigma_1}} = 3,79\%, \overline{\Delta_{\sigma_2}} = 1,84\%, \overline{\Delta_{b_1}} = 15,65\%, \overline{\Delta_{a_2}} = 10,45\%, \overline{\Delta_{\alpha}} = 9,25\%.$$

Для интегрального оценивания качества предложенного алгоритма был использован следующий показатель:

$$\Delta_{integr} = \frac{\sum_i^{Ns} \left(F_{teor}(x, \mu_1, \sigma_1, a_1, b_1, \mu_2, \sigma_2, a_2, b_2, \alpha) - F_{pract}(x, \tilde{\mu}_1, \tilde{\sigma}_1, \tilde{a}_1, \tilde{b}_1, \tilde{\mu}_2, \tilde{\sigma}_2, \tilde{a}_2, \tilde{b}_2, \tilde{\alpha}) \right)^2}{\sum_i^{Ns} F_{teor}(x, \mu_1, \sigma_1, a_1, b_1, \mu_2, \sigma_2, a_2, b_2, \alpha)^2},$$

значения которого в проведенных вычислительных экспериментах изменялись в диапазоне $\left[(\Delta_{integr})_{\min}, (\Delta_{integr})_{\max} \right]$, где $(\Delta_{integr})_{\min} = 0.006$, $(\Delta_{integr})_{\max} = 1.776$. Наилучший результат $(\Delta_{integr} = (\Delta_{integr})_{\min})$ получен при следующих настройках ГА:

$\langle S = \text{турнирная}, M = \text{адаптивная}, C = \text{эвристический}, Cfr = 0.6, P = 30 \rangle$,

наихудший $(\Delta_{integr} = (\Delta_{integr})_{\max})$ –

$\langle S = \text{рулетка}, M = \text{адаптивная}, C = \text{одноточечный}, Cfr = 0.3, P = 15 \rangle$,

Соответствующие параметры распределений представлены в таблице 3.7.

Таблица 3.7. Наилучший и наихудший интегральный показатели

| Параметры | μ_1 | σ_1 | a_1 | \tilde{b}_1 | μ_2 | σ_2 | a_2 | \tilde{b}_2 | α | Δ_{integr} |
|------------------------|---------|------------|-------|---------------|---------|------------|-------|---------------|----------|-------------------|
| Теоретические значения | 7 | 2 | 3 | 14 | 14 | 1 | 6 | 18 | 0.5 | – |
| Наихудший результат | 7.225 | 1.869 | 3.023 | 6.109 | 14.014 | 0.956 | 9.014 | 17.977 | 0.813 | 1.776 |
| Наилучший результат | 7.225 | 2.113 | 3.023 | 15.203 | 14.014 | 1.051 | 5.269 | 17.977 | 0.474 | 0.006 |

Теоретическая и экспериментальная ПР представлены на рисунке 3.14.

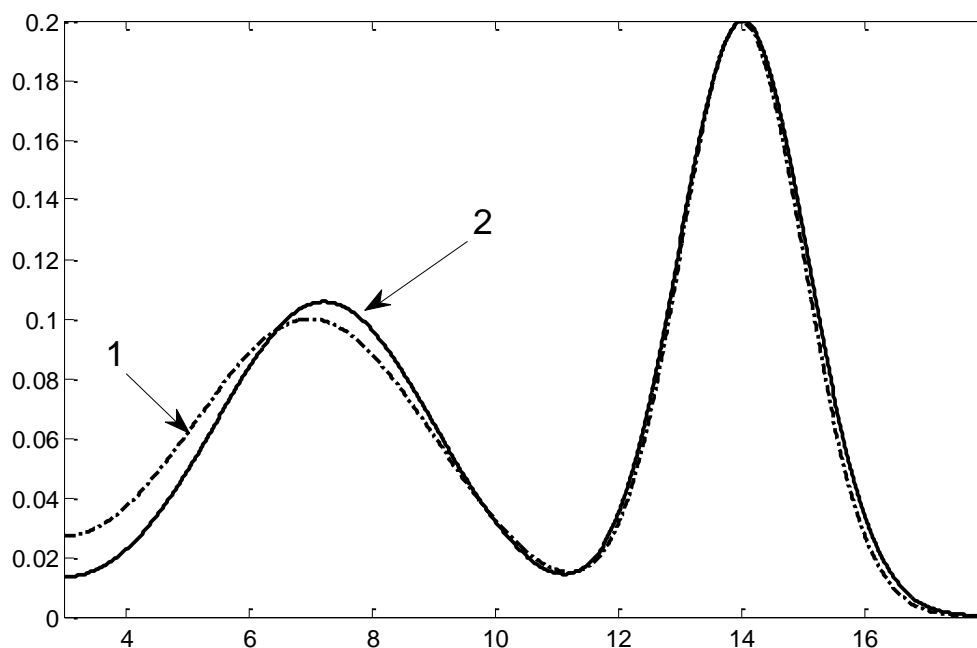


Рисунок 3.14. Теоретическая (1) и экспериментальная ПР (2)

Таким образом, анализ результатов совместного применения аппроксимации Розенблатта-Парзена, метода мнимых источников и ГА в задаче оценивания значений параметров двумодальных ПР СПООР, относящегося к классу 9-ти параметрических распределений, позволяет сделать следующие выводы:

1. При использовании генетических алгоритмов для нахождения одновременно всех параметров изучаемого распределения погрешности оценок превышают 50%, что является следствием высокой размерности решаемой задачи.

2. Предложена методика нахождения параметров двумодальной ПР СПООР, основанный на совместном использовании аппроксимации Розенблатта-Парзена и ГА, и получено подтверждение его работоспособности.

3. Получены оценки точности нахождения параметров двумодальной ПР СПООР, а также интегрального показателя, характеризующего в целом качество оценки ПР СПООР.

3.7 Итерационный алгоритм оценивания параметров случайной величины с двумодальным законом распределения на основе использования аппроксимации Розенблатта-Парзена, метода мнимых источников и ГА

Отметим, что в алгоритме, описанном в предыдущем разделе, удалось уменьшить размерность задачи для ГА, предварительно оценив параметры $\tilde{\mu}_1, \tilde{\mu}_2$ с помощью аппроксимации Розенблатта-Парзена. При этом оказалось,

что точность нахождения ПР в результате уменьшения размерности пространства параметров возрастает. В этой связи понятно, что можно развить подход, использованный в предыдущем разделе, вновь применив ГА, в котором в качестве начальных значений параметров ГА используются уже полученные оценки и т.д.

Таким образом, модифицированная методика оценивания параметров двумодального ПР СПООР реализуется следующей последовательностью действий:

Этап 1. Вычисление в соответствии с методом Розенблатта-Парзена значений функции $f_{RP}(x)$, аппроксимирующей плотность распределения (3.1) случайной величины.

Наличие значений аппроксимирующей функции позволяет оценить примерный вид исходной функции распределения, а также получить оценки координаты левой границы области рассеяния СПООР:

$$\tilde{a}_1 = \min_i (X_i),$$

и правой границы области рассеяния СПООР:

$$\tilde{b}_2 = \max_i (X_i),$$

а также вычислить оценки значений параметров $\tilde{\mu}_1, \tilde{\mu}_2$ – абсцисс локальных максимумов функции $F_{RP}(x)$. (Отметим, что оценки параметров \tilde{a}_1, \tilde{b}_2 на данном этапе оказываются достаточно точными, что позволяет на последующих этапах данного алгоритма считать их известными и не уточнять с помощью ГА.)

Этап 2. Вычисление с помощью ГА значений параметров $\sigma_1, b_1, \sigma_2, a_2, \alpha$. (Здесь параметры μ_1, μ_2 , оценки значений которых были оценены на предыдущем этапе по аппроксимации Розенблатта-Парзена, считаются постоянными.)

Этап 3. Вычисление с помощью ГА значений параметров $\mu_1, b_1, \mu_2, a_2, \alpha$. (Здесь параметры σ_1, σ_2 , оценки значений которых получены на этапе 2, считаются постоянными).

Этап 4. Вычисление с помощью ГА значений параметров $\sigma_1, \sigma_2, \alpha$. (Здесь параметры μ_1, b_1, μ_2, a_2 , оценки значений которых получены на этапе 3, считаются постоянными, в качестве начального приближения параметров $\sigma_1, \sigma_2, \alpha$ используются значения, полученные на этапе 3)

Этап 5. Вычисление, используя ГА, значения параметров μ_1, μ_2 . (Здесь параметры $\sigma_1, b_1, \sigma_2, a_2, \alpha$, оценки значений которых получены на этапе 4, считаются постоянными, в качестве начального приближения параметров μ_1, μ_2 используются значения, полученные на этапе 4)

Этап 6. Вычисление, используя ГА, значения параметров $\sigma_1, \sigma_2, \alpha$. (Здесь параметры μ_1, b_1, μ_2, a_2 , оценки значений которых получены на этапе 5, считаются постоянными, в качестве начального приближения параметров $\sigma_1, \sigma_2, \alpha$ используются значения, полученные на этапе 5).

Здесь в качестве ЦФ можно использовать интегральную погрешность:

$$\Delta_{integr}^{(step)} = \frac{\sum_i^{Ns} \left(F_{RP}(x_i) - F_{pract}^{(step)}(x_i, \tilde{\mu}_1, \tilde{\sigma}_1, \tilde{a}_1, \tilde{b}_1, \tilde{\mu}_2, \tilde{\sigma}_2, \tilde{a}_2, \tilde{b}_2, \tilde{\alpha}) \right)^2}{\sum_i^{Ns} F_{RP}(x_i)^2}, \quad (3.8)$$

где $F_{pract}^{(step)}(x, \tilde{\mu}_1, \tilde{\sigma}_1, \tilde{a}_1, \tilde{b}_1, \tilde{\mu}_2, \tilde{\sigma}_2, \tilde{a}_2, \tilde{b}_2, \tilde{\alpha})$ – ФР СПООР, полученная в результате реализации этапов 1–5 описанного выше алгоритма, $F_{RP}(x)$ – аппроксимация ФР СПООР, вычисленная в соответствии с аппроксимацией Розенблатта-Парзена. Использование данной ЦФ обусловлено тем, что для ее вычисления не требуется значительных вычислительных ресурсов.

Из приведенного выше описания методики оценивания параметров двумодальной ФР СПООР, видно, что на каждом этапе, начиная со второго, происходит уточнение значений параметров распределения, полученных на предыдущем этапе. Для подтверждения достаточности выполнения 6-ти этапов предложенного алгоритма были проведены экспериментальные исследования двумодальных СПООР, параметры которых представлены в таблице 3.8.

Таблица 3.8. Параметры случайных двумодальных распределений

| Параметры распределения | μ_1 | σ_1 | a_1 | b_1 | μ_2 | σ_2 | a_2 | b_2 | α |
|----------------------------|---------|------------|-------|-------|---------|------------|-------|-------|----------|
| 1 | 7 | 2 | 3 | 14 | 14 | 1 | 10 | 18 | 0.5 |
| 2 | 7 | 2 | 3 | 14 | 14 | 1 | 10 | 18 | 0.7 |
| 3 | 4 | 2 | 3 | 14 | 14 | 1 | 10 | 18 | 0.6 |
| 4 | 4 | 2 | 3 | 14 | 17 | 1 | 10 | 18 | 0.5 |
| 5 | 7 | 3 | 3 | 14 | 14 | 1 | 10 | 18 | 0.7 |

Для оценки качества методики использовалось значение погрешности, вычисляемое относительно теоретической функции распределения случайной последовательности:

$$\Delta_{integr}^{(step)} = \frac{\sum_i^{Ns} \left(F_{teor}(x_i, \mu_1, \sigma_1, a_1, b_1, \mu_2, \sigma_2, a_2, b_2, \alpha) - F_{pract}^{(step)}(x_i, \tilde{\mu}_1, \tilde{\sigma}_1, \tilde{a}_1, \tilde{b}_1, \tilde{\mu}_2, \tilde{\sigma}_2, \tilde{a}_2, \tilde{b}_2, \tilde{\alpha}) \right)^2}{\sum_i^{Ns} F_{teor}(x_i, \mu_1, \sigma_1, a_1, b_1, \mu_2, \sigma_2, a_2, b_2, \alpha)^2},$$

где $F_{teor}(x, \mu_1, \sigma_1, a_1, b_1, \mu_2, \sigma_2, a_2, b_2, \alpha)$ – теоретическая функция распределения. Результаты применения итерационного алгоритма к СПООР №1-5 приведены в Приложении Б.

Из рисунков П.Б.1-П.Б.5 видно, что функцию интегральной погрешности можно аппроксимировать функцией вида $f_{integral}(x) = a \cdot e^{-b \cdot x} + c$. Значения коэффициентов a, b, c для исследованных СПООР представлены в таблице 3.9.

Таблица 3.9. Параметры функций, аппроксимирующих поэтапную интегральную погрешность

| № СПООР | a | b | c |
|---------|--------|--------|---------|
| 1 | 0,0642 | 0,2136 | 0,0002 |
| 2 | 0,0371 | 1,9481 | 0,0136 |
| 3 | 0,0245 | 0,5606 | 0,0213 |
| 4 | 0,0396 | 1,1779 | 0,0187 |
| 5 | 0,0654 | 0,1578 | -0,0001 |

Для подтверждения достаточности перечисленных выше этапов алгоритма были проведены вычисления, аналогичные вычислениям, выполняемым на этапе 2. При этом в качестве начальных значений использовались значения параметров, полученные в результате выполнения этапов 1–6 алгоритма. Значения интегральной погрешности ФР, вычисленной дополнительно $f_{integral}^{\{7\}}$, а также модуль разности интегральных погрешностей ПР $f_{integral}^{\{7\}}$ и $f_{integral}^{\{6\}}$ представлены в таблице 3.10.

Таблица 3.10. К сравнению значений интегральных погрешностей

| № СПООР | $f_{integral}^{\{7\}}$ | $ f_{integral}^{\{7\}} - f_{integral}^{\{6\}} $ | $ f_{integral}^{\{7\}} - f_{integral}^{\{6\}} / f_{integral}^{\{7\}}, \%$ |
|---------|------------------------|---|--|
| 1 | 0,0144 | $3,4264 \cdot 10^{-3}$ | 23,7944 |
| 2 | 0,0137 | $2,6731 \cdot 10^{-7}$ | 0,0020 |
| 3 | 0,0218 | $3,6310 \cdot 10^{-4}$ | 1,6656 |
| 4 | 0,0187 | $2,3297 \cdot 10^{-5}$ | 0,1246 |
| 5 | 0,0217 | $3,7045 \cdot 10^{-3}$ | 17,0714 |

Из таблицы 3.10 видно, что на седьмом этапе изменение относительной интегральной погрешности ФР не превосходит 24%. Следовательно, в предложенном итерационном алгоритме, действительно, для нахождения ФР с приемлемой точностью можно ограничиться шестью этапами.

Таким образом, полученные результаты позволяют сделать вывод о том, что предложенная методика в целом показывает результат лучший, чем двухэтапный алгоритм и метод Розенблатта-Парзена. С каждым следующим этапом предложенной методики наблюдается тенденция к уменьшению интегральной погрешности. При этом, начиная с 6 этапа, изменение значения интегральной погрешности становится уже несущественным, поэтому можно утверждать, что для получения оптимального результата вполне достаточно 6 этапов.

3.8 Выводы по главе

1. Дана формулировка задачи оценки параметров двумодальных ФР и ПР в терминах ГА.

2. Получены количественные подтверждения низкой точности оценок двумодальных ФР и ПР в соответствии с методикой, обоснованной в Главе 2.

3. Исследованы особенности численных оценок параметра размытости в аппроксимации Розенблатта-Парзена, результаты которых позволили объяснить немонотонность информационного функционала у ядер с неограниченной областью определения и обосновать границы областей, в которых следует искать оптимальные значения параметров размытости для использованных в работе ядерных функций.

4. Предложена модифицированная формула для вычисления значений информационного функционала целочисленных случайных последовательностей.

5. Проведен сравнительный анализ точности одномодальных ФР и ПР, оцениваемых с помощью ГА, и аппроксимации Розенблатта-Парзена, результаты которого показывают, оценки на основе метода мнимых источников оказываются более точными.

7. Обоснована комплексная методика оценки ФР и ПР СПООР, основанная на совместном использовании метода мнимых источников, ГА и аппроксимации Розенблатта-Парзена и подтверждена ее работоспособность.

Результаты исследования, приведенного в данной главе, опубликованы в [61,62,65,66,68,119].

4 Описание программной библиотеки «ES&RP», обеспечивающей практическое использование математического и алгоритмического обеспечения для оценки ФР и ПВ СПООР

Для практического использования описанного в предыдущих разделах математического и алгоритмического обеспечения для решения задачи оценивания ФР и ПР случайных величин с ООР была разработана программная библиотека, описываемая в данной главе.

4.1 Назначение программной библиотеки

Программная библиотека ES&RP содержит набор функций для решения задачи аппроксимации функций распределений случайных величин с ООР в соответствии с подходами, предложенными в Главе 2 и Главе 3.

Для оценки параметров функций распределения и плотностей распределений в методе мнимых источников в программной библиотеке используются генетические алгоритмы.

Для вычисления аппроксимации Розенблатта-Парзена предусмотрена возможность использования следующих ядерных функций: функция Коши, функция Епанчикова, функция Фишера, функция Лапласа, логистическая функция, нормальная функция, квадратичная функция, треугольная функция, равномерная функция.

В состав библиотеки также включены функции, позволяющие реализовать итерационный алгоритм оценивания параметров функций распределения и плотностей распределения двумодальных случайных величин с ограниченной областью рассеяния, основанный на совместном использовании метода мнимых источников и аппроксимации Розенблатта-Парзена.

4.2 Используемые технические средства

Тип реализующей ЭВМ: IBM-совместимая, класса Pentium и выше

Язык: MATLAB (MATLAB Compiler V. 4.13)

ОС: Windows (Win-32).

Объем программной библиотеки: 45.2 Кб

4.3 Описание модулей программной библиотеки

4.3.1 Описание функции BDconstraints.m

function [A, b, LB, UB] = BDconstraints(a1, a2)

Функция, возвращающая:

A – матрица уравнения линейных ограничений $A \cdot x < b$

b – вектор уравнения линейных ограничений $A \cdot x < b$

LB – вектор нижних границ параметров

UB – вектор верхних границ параметров

Входные параметры

$a1$ – нижняя граница области рассеяния

$a2$ – верхняя граница области рассеяния

4.3.2 Описание функции **BoundDistrFit.m**

function $z = \text{BoundDistrFit}(x, \text{arr})$

Функция, возвращающая:

z – значение функции приспособленности в ГА для метода мнимых источников

Входные параметры

x – вектор параметров ГА

arr – вектор значений случайной последовательности

4.3.3 Описание функции распределения **checkDistrType.m**

function $[\text{type}, h, p] = \text{checkDistrType}(x)$

Функция, возвращающая

type – тип распределения (1 – равномерное, 2 – смешанное, 3 – нормальное)

h – решение по тесту К-С, 1 – отвергается нуль-гипотеза, 0 – принимается

p – р-значение

Входные параметры

x – случайная последовательность

4.3.4 Описание функции **normLADcdf.m**

function $z = \text{normLADcdf}(x, x0, \text{Sigma}, L, N)$

функция, возвращающая значение ФР нормального закона распределения с ограниченной областью рассеяния

Входные параметры

x – вектор-строка

$x0$ – положение действительного центра рассеяния относительно центра области рассеяния

Sigma – стандартное отклонение при отсутствии ограничений

L – размер области рассеяния

N – количество фиктивных источников, расположенных слева/справа от одной из отражающих плоскостей.

N = 1 - 3 функции, N = 2 - 5 функций, и т.д.

4.3.5 Описание функции normLADcdf2.m

function z=normLADcdf2(x,Mu,Sigma,x1,x2,N)

функция, возвращающая значение ФР нормального закона распределения с ограниченной областью рассеяния

Входные параметры

x – вектор-строка

Mu – положение действительного центра рассеяния

Sigma – стандартное отклонение при отсутствии ограничений

x1, x2 – границы области рассеяния

N – количество используемых фиктивных источников, расположенных слева/справа от одной из отражающих плоскостей.

N = 1 – 3 функции, N = 2 – 5 функций, и т.д.

4.3.6 Описание функции normLADinv.m

function z=normLADinv(p,x0,Sigma,L,N)

функция, возвращающая значение квантиля функции распределения нормального закона с ограниченной областью рассеяния

Входные параметры

p – вектор-строка

X0 – положение действительного центра рассеяния относительно центра области рассеяния

Sigma – стандартное отклонение при отсутствии ограничений

L – размер области рассеяния

N – количество фиктивных источников, расположенных слева/справа от одной из отражающих плоскостей.

N = 1 – 3 функции, N = 2 – 5 функций, и т.д.

4.3.7 Описание функции normLADinv2.m

function z=normLADinv2(p,Mu,Sigma,a1,a2,N)

функция, возвращающая значение квантиля функции распределения нормального закона с ограниченной областью рассеяния

Входные параметры

p – вектор-строка

μ – положение действительного центра рассеяния

σ – стандартное отклонение при отсутствии ограничений

a_1, a_2 – границы области рассеяния

N – количество фиктивных источников, расположенных слева/справа от одной из отражающих плоскостей.

$N = 1$ – 3 функции, $N = 2$ – 5 функций, и т.д.

4.3.8 Описание функции **normLADpdf.m**

`function p=normLADpdf(x,x0,Sigma,L,N)`

функция, возвращающая значение плотности нормального закона распределения с ограниченной областью рассеяния

Входные параметры

x – вектор-строка

x_0 – положение действительного центра рассеяния относительно центра области рассеяния

σ – стандартное отклонение при отсутствии ограничений

L – размер области рассеяния

N – количество фиктивных источников, расположенных слева/справа от одной из отражающих плоскостей.

$N = 1$ – 3 функции, $N = 2$ – 5 функций, и т.д.

4.3.9 Описание функции **normLADpdf2.m**

`function p=normLADpdf2(x,Mu,Sigma,x1,x2,N)`

функция, возвращающая значение плотности нормального закона распределения с ограниченной областью рассеяния

Входные параметры

x – вектор-строка

μ – положение действительного центра рассеяния

σ – стандартное отклонение при отсутствии ограничений

x_1, x_2 – границы области рассеяния

N – количество используемых фиктивных источников, расположенных слева/справа от одной из отражающих плоскостей.

$N = 1$ – 3 функции, $N = 2$ – 5 функций, и т.д.

4.3.10 Описание функции solveBoundDistrGA.m

function [Mu, Sigma, a1, a2, y] = solveBoundDistrGA(x, cx)

Функция, возвращающая

Mu, Sigma, a1, a2 – параметры распределения

y – значения функции плотности распределения для значений аргумента cx

Входные параметры

x – случайная последовательность

cx – значения по оси OX для функции плотности

4.3.11 Описание функции doublePeakCon1.m

function [A, b, LB, UB] = doublePeakCon1(zD1, zD2)

Функция, возвращающая

A – матрица уравнения линейных ограничений $A \cdot x < b$

b – вектор уравнения линейных ограничений $A \cdot x < b$

LB – вектор нижних границ параметров

UB -- вектор верхних границ параметров

Функция применяется для первого шага итерационного алгоритма

Входные параметры

zD1 – структура, содержащая оценки параметров первой составляющей

zD2 – структура, содержащая оценки параметров второй составляющей

4.3.12 Описание функции doublePeakCon2.m

function [A, b, LB, UB] = doublePeakCon2(zD1, zD2)

Функция, возвращающая

A – матрица уравнения линейных ограничений $A \cdot x < b$

b – вектор уравнения линейных ограничений $A \cdot x < b$

LB – вектор нижних границ параметров

UB – вектор верхних границ параметров

Функция применяется для второго шага итерационного алгоритма

Входные параметры

zD1 – структура, содержащая оценки параметров первой составляющей

zD2 – структура, содержащая оценки параметров второй составляющей

4.3.13 Описание функции doublePeakCon3.m

function [A, b, LB, UB] = doublePeakCon3(zD1, zD2)

Функция, возвращающая

A – матрица уравнения линейных ограничений $A \cdot x < b$

b – вектор уравнения линейных ограничений $A \cdot x < b$

LB – вектор нижних границ параметров

UB – вектор верхних границ параметров

Функция применяется для третьего шага итерационного алгоритма

Входные параметры

zD1 – структура, содержащая оценки параметров первой составляющей

zD2 – структура, содержащая оценки параметров второй составляющей

4.3.14 Описание функции doublePeakCon4.m

function [A, b, LB, UB] = doublePeakCon4(zD1, zD2)

Функция, возвращающая

A – матрицу уравнения линейных ограничений $A \cdot x < b$

b – вектор уравнения линейных ограничений $A \cdot x < b$

LB – вектор нижних границ параметров

UB – вектор верхних границ параметров

Функция применяется для четвертого шага итерационного алгоритма

Входные параметры

zD1 – структура, содержащая оценки параметров первой составляющей

zD2 – структура, содержащая оценки параметров второй составляющей

4.3.15 Описание функции doublePeakCon5.m

function [A, b, LB, UB] = doublePeakCon5(zD1, zD2, zalpha)

Функция, возвращающая

A – матрицу уравнения линейных ограничений $A \cdot x < b$

b – вектор уравнения линейных ограничений $A \cdot x < b$

LB – вектор нижних границ параметров

UB – вектор верхних границ параметров

Функция применяется для пятого шага итерационного алгоритма

Входные параметры

zD1 – структура, содержащая оценки параметров первой составляющей

zD2 – структура, содержащая оценки параметров второй составляющей

4.3.16 Описание функции **doublePeakFit01.m**

function z = doublePeakFit01(x, zD1, zD2, fPR, cx)

Функция, возвращающая

z – значение функции приспособленности в ГА для двумодального распределения

Функция применяется для первого шага итерационного алгоритма

Входные параметры

x – вектор параметров ГА

zD1 – структура, содержащая вычисленные параметры первой составляющей распределения

zD2 – структура, содержащая вычисленные параметры второй составляющей распределения

fPR – значения аппроксимации Розенблатта-Парзена

cx – значения аргумента для вычисления плотности распределения

4.3.17 Описание функции **doublePeakFit02.m**

function z = doublePeakFit02(x, zD1, zD2, fPR, cx)

Функция, возвращающая

z – значение функции приспособленности в ГА для двумодального распределения

Функция применяется для второго шага итерационного алгоритма

Входные параметры

x – вектор параметров ГА

zD1 – структура, содержащая вычисленные параметры первой составляющей распределения

zD2 – структура, содержащая вычисленные параметры второй составляющей распределения

fPR – значения аппроксимации Розенблатта-Парзена

cx – значения аргумента для вычисления плотности распределения

4.3.18 Описание функции **doublePeakFit03.m**

function z = doublePeakFit03(x, zD1, zD2, fPR, cx)

Функция, возвращающая

z – значение функции приспособленности в ГА для двумодального распределения

Функция применяется для третьего шага итерационного алгоритма

Входные параметры

x – вектор параметров ГА

$zD1$ – структура, содержащая вычисленные параметры первой составляющей распределения

$zD2$ – структура, содержащая вычисленные параметры второй составляющей распределения

fPR – значения аппроксимации Розенблатта-Парзена

cx – значения аргумента для вычисления плотности распределения

4.3.19 Описание функции **doublePeakFit04.m**

`function z = doublePeakFit04(x, zD1, zD2, zalpha, fPR, cx)`

Функция, возвращающая

z – значение функции приспособленности в ГА для двумодального распределения

Функция применяется для четвертого шага итерационного алгоритма

Входные параметры

x – вектор параметров ГА

$zD1$ – структура, содержащая вычисленные параметры первой составляющей распределения

$zD2$ – структура, содержащая вычисленные параметры второй составляющей распределения

fPR – значения аппроксимации Розенблатта-Парзена

$zalpha$ – вычисленное значение параметра α

cx – значения аргумента для вычисления плотности распределения

4.3.20 Описание функции **doublePeakFit05.m**

`function z = doublePeakFit05(x, zD1, zD2, fPR, cx)`

Функция, возвращающая

z – значение функции приспособленности в ГА для двумодального распределения

Функция применяется для пятого шага итерационного алгоритма

Входные параметры

x – вектор параметров ГА

zD1 – структура, содержащая вычисленные параметры первой составляющей распределения

zD2 – структура, содержащая вычисленные параметры второй составляющей распределения

fPR – значения аппроксимации Розенблатта-Парзена

cx – значения аргумента для вычисления плотности распределения

4.3.21 Описание функции doublePeakLADinv2.m

```
function z=doublePeakLADinv2(p,...
    Mu1,Sigma1,a1,b1,...
    Mu2,Sigma2,a2,b2,...
    alpha, N)
```

Функция, возвращающая

z – значение квантиля функции распределения двумодального распределения

Входные параметры

p – вектор значений вероятности

Mu1, Sigma1, a1, b1 – параметры первого распределения

Mu2, Sigma2, a2, b2 – параметры второго распределения

alpha – коэффициент первого распределения

N – количество мнимых источников

4.3.22 Описание функции solveDoublePeak.m

```
function [D11, D12, alpha1, fPract1,...
    D21, D22, alpha2, fPract2,...
    D31, D32, alpha3, fPract3,...
    D41, D42, alpha4, fPract4,...
    D51, D52, alpha5, fPract5, fPR] = solveDoublePeak(arr, cx)
```

Функция, возвращающая

D#1 – структуру, содержащую параметры первого распределения на этапе #

D#2 – структуру, содержащую параметры второго распределения на этапе #

alpha# – коэффициент первого распределения на этапе #

fpract# – вектор значений аппроксимации распределения на этапе #

fPR – значения аппроксимации Розенблатта-Парзена

Входные параметры

cx – значения по оси OX для плотности распределения

arr – случайная последовательность

4.3.23 Описание функции solveDoublePeak1.m

`function [D1, D2, alpha, y] = solveDoublePeak1(cx, zD1, zD2, zalpha, fPR)`

Функция, возвращающая

$D1$ – структуру, содержащую вычисленные параметры первой составляющей распределения

$D2$ – структуру, содержащую вычисленные параметры второй составляющей распределения

α – вычисленное значение параметра α

y – вектор значений аппроксимации плотности распределения

Функция применяется для первого шага итерационного алгоритма

Входные параметры

cx – значения аргумента для вычисления плотности распределения

$zD1$ – структура, содержащая вычисленные параметры первой составляющей распределения на предыдущем этапе

$zD2$ – структура, содержащая вычисленные параметры второй составляющей распределения на предыдущем этапе

fPR – значения аппроксимации Розенблатта-Парзена

$zalpha$ – вычисленное значение параметра α на предыдущем этапе

fPR – значения аппроксимации Розенблатта-Парзена

4.3.24 Описание функции solveDoublePeak2.m

`function [D1, D2, alpha, y] = solveDoublePeak2(cx, zD1, zD2, zalpha, fPR)`

Функция, возвращающая

$D1$ – структуру, содержащую вычисленные параметры первой составляющей распределения

$D2$ – структуру, содержащую вычисленные параметры второй составляющей распределения

α – вычисленное значение параметра α

y – вектор значений аппроксимации плотности распределения

Функция применяется для второго шага итерационного алгоритма

Входные параметры

cx – значения аргумента для вычисления плотности распределения

$zD1$ – структура, содержащая вычисленные параметры первой составляющей распределения на предыдущем этапе

zD2 – структура, содержащая вычисленные параметры второй составляющей распределения на предыдущем этапе

fPR – значения аппроксимации Розенблатта-Парзена

zalpha – вычисленное значение параметра alpha на предыдущем этапе

fPR – значения аппроксимации Розенблатта-Парзена

4.3.25 Описание функции solveDoublePeak3.m

function [D1, D2, alpha, y] = solveDoublePeak3(cx, zD1, zD2, zalpha, fPR)

Функция, возвращающая

D1 – структуру, содержащую вычисленные параметры первой составляющей распределения

D2 – структуру, содержащую вычисленные параметры второй составляющей распределения

alpha – вычисленное значение параметра alpha

y – вектор значений аппроксимации плотности распределения

Функция применяется для третьего шага итерационного алгоритма

Входные параметры

cx – значения аргумента для вычисления плотности распределения

zD1 – структура, содержащая вычисленные параметры первой составляющей распределения на предыдущем этапе

zD2 – структура, содержащая вычисленные параметры второй составляющей распределения на предыдущем этапе

fPR – значения аппроксимации Розенблатта-Парзена

zalpha – вычисленное значение параметра alpha на предыдущем этапе

fPR – значения аппроксимации Розенблатта-Парзена

4.3.26 Описание функции solveDoublePeak4.m

function [D1, D2, alpha, y] = solveDoublePeak4(cx, zD1, zD2, zalpha, fPR)

Функция, возвращающая

D1 – структуру, содержащую вычисленные параметры первой составляющей распределения

D2 – структуру, содержащую вычисленные параметры второй составляющей распределения

alpha – вычисленное значение параметра alpha

y – вектор значений аппроксимации плотности распределения

Функция применяется для четвертого шага итерационного алгоритма

Входные параметры

cx – значения аргумента для вычисления плотности распределения

zD1 – структура, содержащая вычисленные параметры первой составляющей распределения на предыдущем этапе

zD2 – структура, содержащая вычисленные параметры второй составляющей распределения на предыдущем этапе

fPR – значения аппроксимации Розенблатта-Парзена

zalpha – вычисленное значение параметра alpha на предыдущем этапе

fPR – значения аппроксимации Розенблатта-Парзена

4.3.27 Описание функции solveDoublePeak5.m

function [D1, D2, alpha, y] = solveDoublePeak5(cx, zD1, zD2, zalpha, fPR)

Функция, возвращающая

D1 – структуру, содержащую вычисленные параметры первой составляющей распределения

D2 – структуру, содержащую вычисленные параметры второй составляющей распределения

alpha – вычисленное значение параметра alpha

y – вектор значений аппроксимации плотности распределения

Функция применяется для пятого шага итерационного алгоритма

Входные параметры

cx – значения аргумента для вычисления плотности распределения

zD1 – структура, содержащая вычисленные параметры первой составляющей распределения на предыдущем этапе

zD2 – структура, содержащая вычисленные параметры второй составляющей распределения на предыдущем этапе

fPR – значения аппроксимации Розенблатта-Парзена

zalpha – вычисленное значение параметра alpha на предыдущем этапе

fPR – значения аппроксимации Розенблатта-Парзена

4.3.28 Описание ядерной функции Епанчикова kernelEpanchikov.m

function [z, m, p] = kernelEpanchikov(arg)

Функция, возвращающая

z – значение ядерной функции

m – мантисса

p – порядок

Входные параметры

arg – аргумент ядерной функции

4.3.29 Описание ядерной функции Фишера kernelFischer.m

function [z, m, p] = kernelFischer(arg)

Функция, возвращающая

z – значение ядерной функции

m – мантисса

p – порядок

Входные параметры

arg – аргумент ядерной функции

4.3.30 Описание ядерной функции Лапласа kernelLaplace.m

function [z, m, p] = kernelLaplace(arg)

Функция, возвращающая

z – значение ядерной функции

m – мантисса

p – порядок

Входные параметры

arg – аргумент ядерной функции

4.3.31 Описание ядерной логистической функции kernelLogist.m

function [z, m, p] = kernelLogist(arg)

Функция, возвращающая

z – значение ядерной функции

m – мантисса

p – порядок

Входные параметры

arg – аргумент ядерной функции

4.3.32 Описание нормальной ядерной функции kernelNorm.m

function [z, m, p] = kernelNorm(arg)

Функция, возвращающая

z – значение ядерной функции

m – мантисса

p – порядок

Входные параметры

arg – аргумент ядерной функции

4.3.33 Описание нормальной ядерной функции kernelSquare.m

function [z, m, p] = kernelSquare(arg)

Функция, возвращающая

z – значение ядерной функции

m – мантисса

p – порядок

Входные параметры

arg – аргумент ядерной функции

4.3.34 Описание треугольной ядерной функции kernelTriangle.m

function [z, m, p] = kernelTriangle(arg)

Функция, возвращающая

z – значение ядерной функции

m – мантисса

p – порядок

Входные параметры

arg – аргумент ядерной функции

4.3.35 Описание равномерной ядерной функции kernelUniform.m

function [z, m, p] = kernelUniform(arg)

Функция, возвращающая

z – значение ядерной функции

m – мантисса

p – порядок

Входные параметры

arg – аргумент ядерной функции

4.3.36 Описание функции PRconstraints.m

function [A, b, LB, UB] = PRconstraints()

Функция, возвращающая

A – матрицу уравнения линейных ограничений $A \cdot x < b$

b – вектор уравнения линейных ограничений $A \cdot x < b$

LB – вектор нижних границ параметров

UB – вектор верхних границ параметров

Входные параметры

4.3.37 Описание функции PRdensity.m

function z = PRdensity(x, arr, h, fnc)

Функция, возвращающая

z – значение аппроксимации Розенблатта-Парзена

Входные параметры

x – вектор значений аргумента для вычисления аппроксимации

arr – случайная последовательность

h – значение параметра размытости

fnc – ссылка на ядерную функцию

4.3.38 Описание функции PRFindOptimH.m

function h = PRFindOptimH(arr, varargin)

Функция, возвращающая

h – вектор оптимальных значений параметра размытости для каждой ядерной функции из списка

Входные параметры

arr – случайная последовательность

varargin – список имен ядерных функций

4.3.39 Описание функции PRfunctional.m

function z = PRfunctional(arr, h, fnc)

Функция, возвращающая

z – значение информационного функционала

Входные параметры

h – значение параметра размытости

arr – случайная последовательность

fnc – ссылка на ядерную функцию

4.3.40 Описание функции solvePR.m

function [y, kernel, h, J] = solvePR(arr, cx)

Функция, возвращающая

y – значение аппроксимации Розенблатта-Парзена

kernel – наименование наилучшей ядерной функции

h – оптимальное значение параметра размытости

J – оптимальное значение информационного функционала

Входные параметры

arr – случайная последовательность

cx – значения аргумента для вычисления плотности распределения

4.3.41 Описание функции **DistributionInv.m**

function z = DistributionInv(x, F, N)

Функция, возвращающая

z – случайную последовательность с заданным законом распределения

Входные параметры

x – значения по оси ОХ

F – значения функции распределения

N – число элементов случайной последовательности

4.3.42 Описание функции интегральной ошибки **integralError.m**

function z = integralError(fTeor, fPract)

Функция, возвращающая

z - значение интегральной ошибки

Входные параметры

fTeor - теоретическая плотность распределения

fPract - практическая плотность распределения

4.4 Примеры использования программной библиотеки

4.4.1 Пример использования программной библиотеки для оценивания одномодального распределения

Постановка задачи: имеется случайная последовательность *Arr*. Предполагая, что распределение случайной величины подчиняется одномодальному нормальному закону в ООР, требуется определить параметры распределения σ, μ, a, b , где

μ – математическое ожидание;

σ – СКО;

a, b – границы области рассеяния.

Задаем параметры исходного распределения

```
clc; close all; clear;
Mu = 100;
Sigma = 40;
a = 0;
b = 160;
```

Генерируем тестовую случайную последовательность

```
numElements = 250;
Arr = GenerateBoundData(Mu, Sigma, a, b, numElements);
```

Визуализируем эмпирическую ПР

```
xmin = min(Arr);
xmax = max(Arr);
dx = (xmax - xmin) / 1001;
x = xmin:dx:xmax;
nBins = 10;
[cx, ~, cyDensity] = MakeBar(Arr, nBins);
figure;
bar(cx, cyDensity, 'FaceColor', [.8 .8 1], 'EdgeColor', [.5 .5 .5]);
```

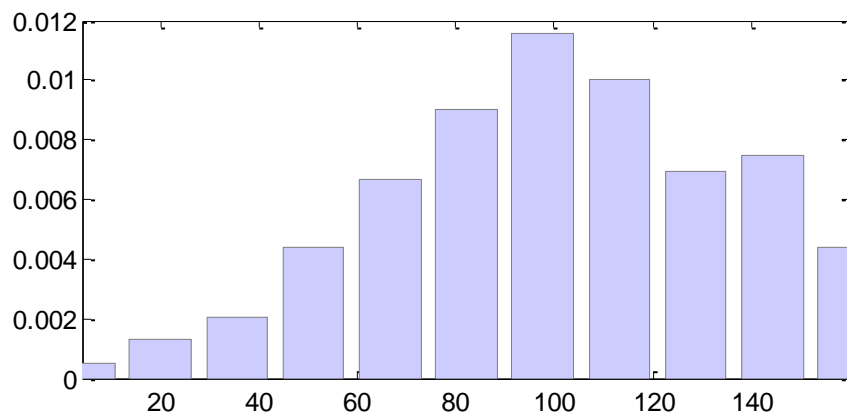


Рисунок 4.1. Эмпирическая плотность распределения

Вызываем библиотечную функцию для вычисления результата

```
[zMu, zSigma, za, zb, y] = solveBoundDistrGA(Arr, x);
```

Визуализируем результат

```
fTeor = normLADpdf2(x, Mu, Sigma, a, b, 5);
fPract = normLADpdf2(x, zMu, zSigma, za, zb, 5);
figure;
bar(cx, cyDensity, 'FaceColor', [.8 .8 1], 'EdgeColor', [.5 .5 .5]);
hold on;
h1 = plot(x, fTeor);
set(h1, 'Color', 'black', 'LineWidth', 2, 'LineStyle', '-');
h1 = plot(x, fPract);
set(h1, 'Color', 'black', 'LineWidth', 2, 'LineStyle', '-');
xlim([xmin, xmax]);
fprintf('Mu=%0.3f, Sigma= %0.3f, a=%0.3f, b=%0.3f', zMu, zSigma, za, zb);
```

Mu=101.745, Sigma=38.986, a=4.240, b=159.920

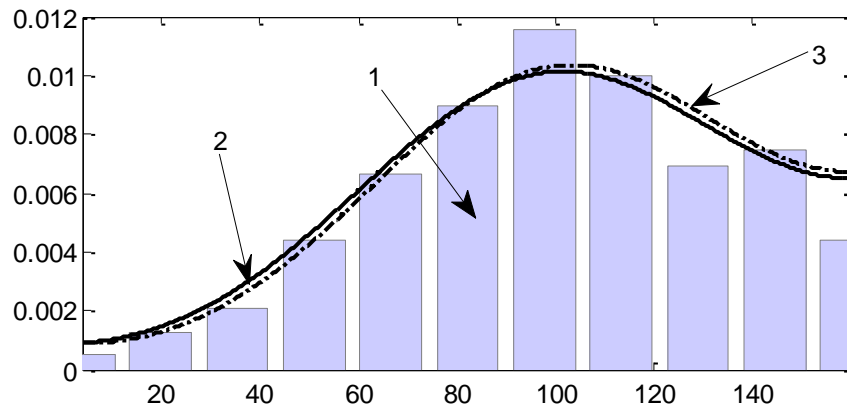


Рисунок 4.2. 1 – Эмпирическая ПР, 2 – теоретическая ПР, 3 – расчетная ПР

Таким образом, заданные и вычисленные параметры распределения и СП с ООР оказываются согласованными друг с другом.

4.4.2 Пример использования программной библиотеки для оценивания двумодального распределения

Постановка задачи: имеется случайная последовательность Arr . Предполагая, что распределение случайной величины подчиняется двумодальному закону в ООР, требуется определить параметры распределения $\sigma_1, \mu_1, a_1, b_1, \sigma_2, \mu_2, a_2, b_2, \alpha$, где

μ_1 – математическое ожидание первой составляющей;

σ_1 – СКО первой составляющей;

a_1, b_1 – границы области рассеяния первой составляющей;

μ_2 – математическое ожидание второй составляющей;

σ_2 – СКО второй составляющей;

a_2, b_2 – границы области рассеяния второй составляющей;

α – доля первой составляющей в общем распределении.

Задаем параметры исходного распределения

```
clc; close all; clear;
format compact;
zMu1 = 10;
zSigma1 = 3;
za1 = 6;
zb1 = 20;
zMu2 = 20;
zSigma2 = 2;
za2 = 10;
zb2 = 30;
zalpha = 0.5;
```

Генерируем тестовую случайную последовательность

```
numElements = 250;
arr = GenerateDoublePeakData(zMu1, zSigma1, za1, zb1, ...
                             zMu2, zSigma2, za2, zb2, zalpha, numElements);
```

Визуализируем эмпирическую ПР

```
xmin = min(arr);
xmax = max(arr);
dx = (xmax - xmin) / 1000;
x = xmin:dx:xmax;
nBins = 12;
[cx, ~, cyDensity] = MakeBar(arr, nBins);
bar(cx, cyDensity, 'FaceColor', [.8 .8 1], 'EdgeColor', [.5 .5 .5]);
xlim([xmin, xmax]);
```

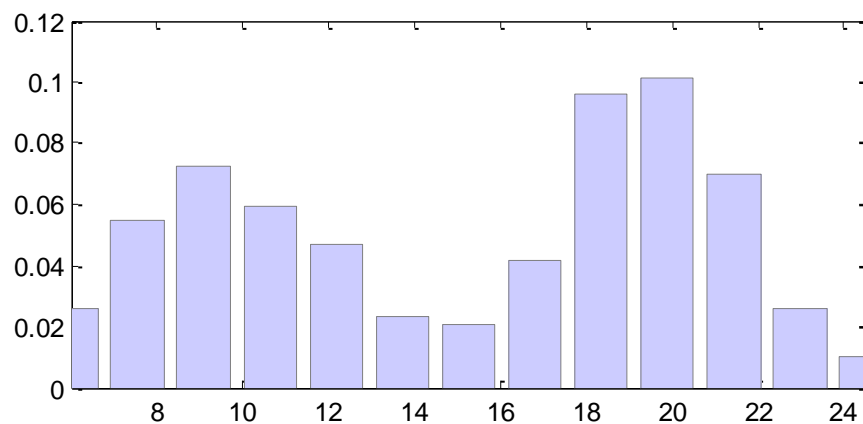


Рисунок 4.3. Эмпирическая плотность распределения

Вызываем библиотечную функцию для вычисления результата

```
[D11, D12, alpha1, fPract1, ...
 D21, D22, alpha2, fPract2, ...
 D31, D32, alpha3, fPract3, ...
 D41, D42, alpha4, fPract4, ...
 D51, D52, alpha5, fPract5, fPR] = solveDoublePeak(arr, x);
```

```
Mu1 = D51.Mu;
Sigma1 = D51.Sigma;
a1 = D51.a;
b1 = D51.b;
Mu2 = D52.Mu;
Sigma2 = D52.Sigma;
a2 = D52.a;
b2 = D52.b;
alpha = alpha5;
```

Визуализируем результат

```
fTeor1 = normLADpdf2(x, zMu1, zSigma1, za1, zb1, 5) * zalpha;
fTeor2 = normLADpdf2(x, zMu2, zSigma2, za2, zb2, 5) * (1 - zalpha);
fTeor = fTeor1 + fTeor2;
fExp = fPract5;
figure;
bar(cx, cyDensity, 'FaceColor', [.8 .8 1], 'EdgeColor', [.5 .5 .5]);
hold on; h1 = plot(x, fTeor);
```

```

set(h1, 'Color', 'black', 'LineWidth', 2, 'LineStyle', '-');
hold on; h1 = plot(x, fExp);
set(h1, 'Color', 'black', 'LineWidth', 2, 'LineStyle', '-');
xlim([xmin, xmax]);
fprintf('Result:\n');
fprintf('Mu1=%0.5f Sigma1=%0.5f a1=%0.5f b1=%0.5f\n', ...
        Mu1, Sigma1, a1, b1);
fprintf('Mu2=%0.5f Sigma2=%0.5f a2=%0.5f b2=%0.5f\n', ...
        Mu2, Sigma2, a2, b2);
fprintf('Alpha=%0.5f \n', alpha);

```

Result:

Mu1=10.04320 Sigma1=2.74039 a1=6.01200 b1=18.57339

Mu2=19.44266 Sigma2=2.64213 a2=12.74516 b2=24.51600

Alpha=0.50453

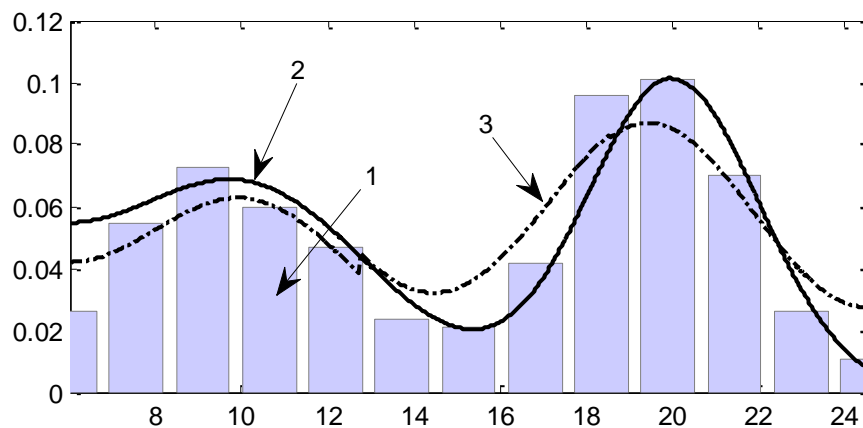


Рисунок 4.4. 1 – Эмпирическая ПР, 2 – теоретическая ПР, 3 – расчетная ПР

Таким образом, заданные и вычисленные параметры распределения и СП с ООР оказываются согласованными друг с другом.

4.5 Выводы по главе

1. Получено подтверждение работоспособности разработанной программной библиотеки в части функций, используемых для оценивания параметров одномодальных распределений с ООР.

2. Получено подтверждение работоспособности разработанной программной библиотеки в части функций, используемых для оценивания параметров двумодальных распределений с ООР.

Заключение

1. Анализ литературы, посвященной проблемам распределений случайных величин с ограниченной областью рассеяния, показал, что существует два альтернативных подхода к построению математических моделей случайных величин с ограниченной областью рассеяния: подход, основанный на описании распределения случайной величины с помощью усеченных функций распределения; подход, базирующийся на работах А. Эйнштейна и М. Смолуховского по теории броуновского движения, в соответствие с которым модель статистического распределения аппроксимируется распределением конечного состояния некоторого случайного процесса без последдействия с ограниченной областью рассеяния.

2. Проведен анализ особенностей УНР и нормального распределения с ограниченной областью рассеяния, позволивший выявить отличия между усеченным нормальным распределением и нормальным распределением с ограниченной областью рассеяния.

3. Проведено исследование особенностей одномодальных распределений случайных последовательностей с ограниченной областью рассеяния на основе статистического моделирования, результаты которого позволили обосновать рекомендации по выбору параметров ГА.

4. Разработаны модификации алгоритмов нахождения оптимального значения параметра размытости h^* при использовании аппроксимации Розенблатта-Парзена, в котором учитывается особенность машинного нуля ПК; нахождения оптимального значения параметра размытости h^* при аппроксимации Розенблатта-Парзена, адаптированный для дискретных целочисленных случайных последовательностей.

5. Разработано математическое и алгоритмическое обеспечение для оценивания функций распределений случайных величин с ограниченной областью рассеяния, включающее в себя:

- алгоритм оценивания параметров одномодальных распределений с ограниченной областью рассеяния на основе ГА;
- алгоритм оценивания параметров двумодальных случайных величин с ограниченной областью рассеяния на основе ГА и аппроксимации Розенблатта-Парзена.

6. Разработана программная библиотека, обеспечивающая практическое использование математического и алгоритмического обеспечения для оценивания функций распределений случайных величин с ограниченной областью рассеяния.

7. Приведены примеры, подтверждающие работоспособность созданной программной библиотеки, при оценивании параметров одномодальных и двумодальных случайных последовательностей с ограниченной областью рассеяния.

Список литературы

1. Айвазян С.А. Прикладная статистика: Основы моделирования и первичная обработка данных: Справ. изд. [Текст] / С. А. Айвазян, И. С. Енюхов, Д. Ж. Мешалкин // М.: Финансы и статистика, 1983. 471 с.
2. Амосов А. А., Дубинский Ю. А., Копченкова Н. П. Вычислительные методы для инженеров. — М.: Мир, 1998.
3. Андерсон Т. Статистический анализ временных рядов. Пер. с англ. [Текст] / Т. Андерсон // М.: Мир, 1976. - 757 с.
4. Барлоу Р., Прошан Ф. Статистическая теория надежности и испытания на безотказность. - М.: Наука, 1984. - 328 с.
5. Бахвалов Н. С. Численные методы (анализ, алгебра, обыкновенные дифференциальные уравнения) - М.: «Наука», 1975. — 632 с.
6. Бахвалов Н.С., Жидков Н.П., Кобельков Г.М. Численные методы. Бином. Лаборатория знаний. 2003. — 640 с.
7. Бернштейн С.Н. Теория вероятностей, 4 изд., М. — Л., 1948.
8. Боровков А.А. Математическая статистика. — М.: Наука, 1984, 472 с.
9. Бородачев Н.А. Точность производства в машиностроении и приборостроении [Текст]/ Н.А. Бородачев, Р.М. Абрашитов, И.М. Веселова// М.: Машиностроение, 1973. —567 с.
10. Бриллинджер Д. Временные ряды: обработка и теория. [Текст] / Д. Бриллинджер // М.: Мир, 1980. -536 с.
11. Ван дер Варден Б.Л. Математическая статистика. [Текст] / Б. Л. Ван дер Варден // М.: Иностранная литература, 1960. - 435 с.
12. Вентцель Е.С. Теория вероятностей. -М.:1969.
13. Вержбицкий В.М. Основы численных методов. Учеб. пособие для вузов. - М.: Высшая школа, 2002.
14. Волков Е. А. Численные методы. — М.: Физматлит, 2003.
15. Гаусс К. Ф. Избранные геодезические сочинения. Т. 1. Метод наименьших квадратов. — М.: Изд-во геодезической литературы, 1957. — 234 с.
16. Гладков Л.А., Курейчик В.В., Курейчик В.М. Генетические алгоритмы / Под ред. В.М. Курейчика. — 2-е изд., исправл. и доп. — М.: ФИЗМАТЛИТ, 2010. — 368 с.
17. Гмурман В. Е. Теория вероятностей и математическая статистика - М., Высш.шк., 2003.- 479 с.
18. Гнеденко Б. В. К истории основных понятий теории вероятностей // История и методология естественных наук. — М.: Изд. МГУ, 1986. —Вып. XXXII. Математика, механика. — С. 81—88.

19. Гнеденко Б. В. Очерк по истории теории вероятностей // Курс теории вероятностей. 8-е изд. — М.: Едиториал УРСС, 2005. — 448 с. — ISBN 5-354-01091-8. — С. 366—435.
20. Гнеденко Б.В. Курс теории вероятностей —М.: Наука. Главное изд-во. Физ.-мат. литературы, 1988. —488 с.
21. Гнеденко Б.В., Беляев Ю.К., Соловьев А.Д. Математические методы в теории надежности. - М.: Наука, 1965. - 524 с.
22. Григорьев А.В. Генетические алгоритмы оптимизации многомерных, многокритериальных задач с нелинейной целевой функцией // Материалы междунар. конф. студ. и асп. по фундаментальным наукам "Ломоносов-2002". Секция "Вычислительная математика и кибернетика". - М.: МАКС-Пресс, 2002.
23. Гулд Х., Тобочник Я. Компьютерное моделирование в физике. В 2-х тт. [Текст]/ Х. Гулд, Я. Тобочник// —М. Мир, 1990. —Т. 2. —399 с.
24. Денисенко В.В. Современные математические методы моделирования сетей передачи данных // Актуальные направления научных исследований XXI века: теория и практика. 2014. № 5–2. С. 178–181.
25. Елисеева И.И. Общая теория статистики. [Текст] / И. И. Елисеева, М. М. Юзбашев // М.: Финансы и статистика, 2004.-656 с.
26. Заляжных В.В., Коптелов А.Е. Статистические методы контроля и управления качеством: Учебное пособие. — Архангельск: Изд-во Архангельского государственного технического университета, 2004. - 88 с.
27. Калиткин Н.Н. Численные методы. — М.: Наука, 1978.
28. Кендалл М. Дж. Временные ряды. [Текст] / М. Дж. Кендалл // М.: Финансы и статистика, 1981. — 199 с.
29. Кендалл М. Дж. Многомерный статистический анализ и временные ряды. [Текст] / М. Дж. Кендалл, А. Стьюарт // М.: Наука, 1976. — 736 с.
30. Киреев В.И., Пантелеев А.В. Численные методы в примерах и задачах — М.: Высш. шк. , 2008, 480 с.
31. Кобзарь, А.И. Прикладная математическая статистика. — М.: Физматлит, 2006.
32. Колесников А.В., Иванов И.П., Басараб М.А. Нелинейно-динамические модели сетевого трафика // Нелинейный мир. 2014. Т. 12, № 4. С. 44–56.
33. Колмогоров А.Н. Основные понятия теории вероятностей. [Текст] / А.Н. Колмогоров // М.: Наука, 1974. - 120 с.
34. Колмогоров А.Н. Теория вероятностей и математическая статистика: Сб. статей. — М.: Наука, 1986. — 535 с.
35. Крамер Г. Математические методы статистики. М.: Мир, 1975. 648 с.

36. Кушнаренко В.М., Зинченко К.А. Техническое диагностирование объектов нефтегазодобычи/ под ред. А.И. Владимирова, В.Я. Кершенбаума: учебное пособие. – М.: Национальный институт нефти и газа, 2005 – 72с.
37. Лаплас П. С. Опыт философии теории вероятностей. 2-е изд. — М.: URSS, 2011. — 208 с.
38. Ляпунов А. М. Собрание сочинений. М.; Л.: Изд-во АН СССР, 1954—1959 Т.1. М.,1954. — 446 с.
39. Магнус Я. Р., Катышев П. К., Пересецкий А. А. Эконометрика. Начальный курс. — М.: Дело, 2007. — 504 с.
40. Майстров Л. Е. Теория вероятностей. Исторический очерк. — М.: Наука, 1967. — 321 с.
41. Маликов И.М., Половко А.М., Романов Н.А., Чукреев П.А. Основы теории и расчёта надёжности. Изд. 2-е, доп. — Л.: Судпромгиз, 1960. - 144с.
42. Марчук Г. И. Методы вычислительной математики. М., Наука, 1977, 456 с.
43. Матвиевский В.Р. Надёжность технических систем [Текст]: учебное пособие/ В.Р. Матвиевский// –М: Московский государственный институт электроники и математики, 2002. 113 с.
44. Мэтьюз Д. Численные методы. Использование MATLAB: учебное издание / Д. Мэтьюз, К. Финк ; пер. с англ. Л.Ф. Козаченко ; под. ред. Ю.В. Козаченко. – М. : Изд. дом Вильямс, 2001. – 720 с.
45. Надёжность магистральных нефтепроводов. Темат. сб. / ВНИИСПТнефть/ - Уфа, 1980, 111 с.
46. Никульчев Е.В., Паяин С.В., Плужник Е.В., Питиков Д.А. Вычисление характеристик динамического хаоса по трафику компьютерных сетей // Фундаментальные исследования. 2014. 8–4. С. 812–816.
47. Орлов А.И. О критериях Колмогорова и Смирнова // Заводская лаборатория. 1995. Т.61. №7. С.59-61.
48. Орлов А.И. Прикладная статистика. [Текст] / А. И. Орлов// М.: Издательство «Экзамен», 2006. - 672 с.
49. Орлов А.И. Распространенная ошибка при использовании критериев Колмогорова и омега-квадрат // Заводская лаборатория.1985. Т.51. №1. С.60-62.
50. Острейковский В.А. Теория надёжности: Учебник для вузов по направлениям "Техника и технологии" и "Технические науки" / В. А. Острейковский . – 2-е изд., испр. - М.: Высш. шк. , 2008. — 464 с.
51. Письменный Д.Т. Конспект лекций по теории вероятностей и математической статистике. - М.: Айрис-пресс, 2004. — 256 с.
52. Половко А.М., Гуров С.В. Основы теории надёжности. — СПб.:БХВ-Петербург, 2006. - 702с

53. Половко А.М., Гуров С.В. Основы теории надёжности. Практикум — СПб.:БХВ-Петербург, 2006. - 560с

54. Поршнев С.В. Теория и алгоритмы аппроксимации эмпирических зависимостей и распределений [Текст]/ С.В. Поршнев, Е.В. Овечкина, В.Е. Каплан// –Екатеринбург: УрО РАН, 2006. –166 с.

55. Поршнев С.В., Божалкин Д.А. Семантический анализатор дампов трафика информационных потоков в компьютерных сетях/ Поршнев С.В., Божалкин Д.А// Свидетельство о государственной регистрации программы для ЭВМ № 2015611426 (Заявка № 2014662922. Дата поступления 12 декабря 2014 г. Дата государственной регистрации в Реестре программ для ЭВМ 29 января 2015 г.).

56. Поршнев С.В., Божалкин Д.А. Технология семантического анализа дампа трафика информационных потоков в компьютерных сетях/ Поршнев С.В., Божалкин Д.А// Информационные технологии. 2014. №11. С. 12-19.

57. Поршнев С.В., Божалкин Д.А. Копосов А.С. Исследование особенностей потоков сетевого трафика в магистральном интернет-канале / Поршнев С.В., Божалкин Д.А., Копосов А.С.//Электросвязь. 2016 №2.С27-33.

58. Поршнев С.В., Божалкин Д.А. Копосов А.С. Опыт использования суперкомпьютера для обработки дампов сетевого трафика магистрального интернет-канала / Поршнев С.В., Божалкин Д.А., Копосов А.С.// Информационные технологии. 2016 №1(22). С 42-47.

59. Поршнев С.В., Гребенкин М.К. Модифицированная жидкостная модель магистрального Интернет-канала / С.В. Поршнев, М.К. Гребенкин // Saarbrücken: LAMBERT Academic Publishing, 2012. 172 с.

60. Поршнев С.В., Копосов А.С. Аналитическое исследование особенностей случайных блужданий броуновской частицы в ограниченной области рассеяния // Фундаментальные исследования. – 2013. – № 4 (часть 1). – стр. 57-64;

61. Поршнев С.В., Копосов А.С. Использование аппроксимации Розенблатта-Парзена для восстановления функции распределения дискретной случайной величины // В мире научных открытий. 2013 №10(46)

62. Поршнев С.В., Копосов А.С. Использование аппроксимации Розенблатта-Парзена для восстановления функции распределения непрерывной случайной величины с ограниченным одномодальным законом распределения // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. Краснодар: КубГАУ, 2013. №08(092). IDA [article ID]: 0921308076. Режим доступа: <http://ej.kubagro.ru/2013/08/pdf/76.pdf>

63. Поршнев С.В., Копосов А.С. Исследование особенностей случайных блужданий броуновской частицы в ограниченной области рассеяния на основе статистического моделирования // Фундаментальные исследования. 2013. №6 (часть 2). С. 284-290

64. Поршнева С.В., Копосов А.С. Методика использования генетических алгоритмов в задаче оценки параметров распределений с ограниченной областью рассеяния // Современные проблемы науки и образования. 2014. №4. С. 168.

65. Поршнева С.В., Копосов А.С. Методика оценивания параметров случайной величины со смешанным двумодальным законом распределения на основе итерационного использования аппроксимации Розенблатта-Парзена и метода мнимых источников // Фундаментальные исследования. – 2014. – № 5 (часть 5). – стр. 974-983;

66. Поршнева С.В., Копосов А.С. Методика оценивания параметров случайной величины со смешанным двумодальным законом распределения на основе совместного использования аппроксимации Розенблатта-Парзена, метода мнимых источников и генетических алгоритмов // Фундаментальные исследования. – 2014. – № 8 (часть 3). – стр. 583-589;

67. Поршнева С.В., Копосов А.С. О выборе математических моделей распределений ограниченных случайных последовательностей [Электр.] / С.В. Поршнева, А.С. Копосов// Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2012. – №10(84). – Режим доступа: <http://ej.kubagro.ru/2012/10/pdf/53.pdf>

68. Поршнева С.В., Копосов А.С. Об особенностях численных оценок параметра размытости плотностей распределений случайных последовательностей в аппроксимации Розенблатта-Парзена // В мире научных открытий. – 2013 №92(08).

69. Рейф Ф. Статистическая физика [Текст]/ Ф. Рейф// –М.: Наука, 1972. –351 С.

70. Рутковская Д., Пилиньский М., Рутковский Л. Нейронные сети, генетические алгоритмы и нечеткие системы: Пер. с польск. И. Д. Рудинского. - М.: Горячая линия -Телеком, 2006. - 452 с.

71. Самарский А. А. Введение в численные методы. Учебное пособие для вузов. 3-е изд., стер. — СПб.: Издательство «Лань», 2005. — 288 с

72. Симахин В.А. Робастные непараметрические оценки: адаптивные оценки взвешенного максимального правдоподобия в условиях статистической априорной неопределенности/ В.А. Симахин // Saarbrücken, Germany: LAP LAMBERT Academic Publishing GmbH & Co. KG, 2011. 292 с.

73. Смирнов Н.В. Теория вероятностей и математическая статистика. Избранные труды. –М.: Наука, 1970. – 289 с.

74. Сызранцев В.Н. Расчет прочностной надежности изделий на основе методов непараметрической статистики / В.Н. Сызранцев, Я.П. Невелев, С.Л. Голофах // Новосибирск: Наука, 2008. 218 с.

75. Тарасенко Ф.П. Непараметрическая статистика. [Текст] / Ф. П. Тарасенко // Томск: Изд-во Томского гос. ун-та, 1976. -292 с.

76. Турчак Л.И. Основы численных методов: учеб. пособие / Л.И. Турчак, П.В. Плотников. – М. : ФИЗМАТЛИТ, 2003. – 304 с.
77. Уилкс С. Математическая статистика. [Текст] / С. Уилкс // М.: Наука, 1967. 574 с.
78. Формалев В.Ф. Численные методы: учебник / В.Ф. Формалев, Д.Л. Ревизников ; под ред. А.И. Кибзуна. – М. : ФИЗМАТЛИТ, 2004. – 400 с.
79. Холлендер М. Непараметрические методы статистики. [Текст] / М. Холлендер // М.: Финансы и статистика, 1983. -518 с.
80. Чебышёв П. Л. Теория вероятностей. Лекции акад. П. Л. Чебышёва, читанные в 1879/80 г.. — М.—Л.: Изд-во АН СССР, 1936. — 253 с.
81. Шейнин О. Б. Теория вероятностей до П. Л. Чебышёва // Историко-математические исследования. — М.: Наука, 1978. — № 23. — С. 284—306.
82. Эйнштейн А. О движении взвешенных в покоящейся жидкости частиц, требуемом молекулярно-кинетической теории теплоты [Текст]// — Ленинград: ОНТИ - Главная редакция общетехнической литературы, 1936. — С. 13–25.
83. Эйнштейн А., Смолуховский М. Брауновское движение [Текст]: сб. статей. — Ленинград: ОНТИ - Главная редакция общетехнической литературы, 1936. –606 с.
84. A. H. Sherman, On Newton-iterative methods for the solution of systems of nonlinear equations, SIAM J. Numer. Anal., 14 (1978), pp. 755–774.
85. Abdullah Konak, David W. Coit, Alice E. Smith, Multi-objective optimization using genetic algorithms: A tutorial // Reliability Engineering and System Safety 91(2006) 992-1007.
86. Agresti, A. Categorical Data Analysis. — Hoboken: John Wiley & Sons, 2013
87. Ashman K.M., Bird C.M. and Zepf S.E. 1994. Detecting bimodality in astronomical datasets. Astronomical Journal 108: 2348
88. Baker, J. E. Adaptive Selection Methods for Genetic Algorithms. Proceedings of an International Conference on Genetic Algorithms and their Application, pp. 101-111, Hillsdale, New Jersey, USA: Lawrence Erlbaum Associates, 1985
89. Booker, L. (1987). Improving search in genetic algorithms. In Genetic Algorithms and Simulating Annealing, ed. L. Davis, pp. 61–73. Morgan Kaufmann, San Mateo, CA.
90. Box, George; Jenkins, Gwilym Time series analysis: forecasting and control, rev. ed. // Oakland, California: Holden-Day. — 1976.
91. Brindle M., Genetic Algorithms for Function Optimization, Ph. Dissertation, University of Alberta, 1981.
92. Cisco Systems. Interconnecting Cisco networking devices // Cisco press. Vol. 1, 2013. 278 p.

93. Darrel Whitley. A Genetic Algorithm Tutorial; November 10, 1993; Technical Report CS-93-103 (Revised); Department of Computer Science, Colorado State University, Fort Collins, US
94. Darrel Whitley. An executable model of a simple genetic algorithm. In Foundations of Genetic Algorithms 2, 1993, ed. D. Whitley. Morgan Kaufmann, San Mateo, CA.
95. David E. Goldberg. Genetic algorithms in search, optimization, and machine learning. Addison-Wesley Publishing Co., Inc., 1989.
96. Davis L., Handbook of Genetic Algorithms, Van Nostrand Reinhold, NY, 1991.
97. Dickey D. A. and Fuller W. A. Distribution of the Estimators for Autoregressive Time Series with a Unit Root / Journal of the American Statistical Association. — 74. — 1979. — p. 427—431.
98. Eyre-Walker A., Keightley P. D. 2007 The distribution of fitness effects of new mutations. Nature Rev. Genet. 8, 610–618.
99. Fonseca, C. M. and Fleming P. J. Genetic Algorithms for Multiple Objective Optimization: Formulation, Discussion and Generalization. Proceedings of the Fifth International Conference on Genetic Algorithms and their Application, pp. 416-423, San Mateo, California, USA: Morgan Kaufmann Publishers, 1993
100. Fonseca, C. M. Multiobjective Genetic Algorithms with Application to Control Engineering Problems. Ph.D. Thesis, Department of Automatic Control and Systems Engineering, University of Sheffield, Sheffield, U.K., 1995.
101. Forrest, Stephanie. Genetic algorithms: principles of natural selection applied to computation. Science, vol.261, p.872-878 (1993).
102. George W. Collins, II. Fundamental Numerical Methods and Data Analysis - 2003
103. Helguerro F. (1904), "Sui Massimi Delle Curve Dimorfiche," Biometrika, 3, 85-98.
104. Hietpas, R.T., Jensen, J.D. & Bolon, D.N. Experimental illumination of a fitness landscape. Proc. Natl. Acad. Sci. USA 108, 7896–7901 (2011)
105. Holzmann H, Vollmer S (2008) A likelihood ratio test for bimodality in two-component mixtures with application to regional income distribution in the EU. AStA Adv Stat Anal 92: 57–69
106. Internet assigned numbers authority: [Электронный ресурс]. URL: <http://www.iana.org/protocols#P> (дата обращения: 01.02.2015).
107. Iversen, G. R., and Gergen, M. (1997), Statistics, the Conceptual Approach, New York: Springer
108. J. F. Traub, Iterative Methods for the Solution of Equations, Prentice Hall, Englewood Cliffs, NJ, 1964.
109. J. M. Ortega and W. C. Rheinboldt, Iterative Solutions of Nonlinear Equations in Several Variables, Academic Press, New York, 1970.

110. K. E. Brenan, S. L. Campbell, and L. R. Petzold, Numerical Solution of Initial Value Problems in Differential-Algebraic Equations, no. 14 in Classics in Applied Mathematics, SIAM, Philadelphia, 1996.
111. Kim, T.-H. and White, H. (2004). On more robust estimation of skewness and kurtosis. *Finance Research Letters* 1, 56-73.
112. L. K. Schubert, Modification of a quasi-Newton method for nonlinear equations with sparse Jacobian, *Math. Comput.*, 24 (1970), pp. 27–30.
113. Mark F. Schilling, Ann E. Watkins, William Watkins. Is Human Height Bimodal? *The American Statistician*, Vol. 56, No. 3, (Aug., 2002), pp. 223-229
114. MAWI Working Group Traffic Archive: [Электронный ресурс]. URL: <http://mawi.nyu.edu/mawi> (дата обращения: 01.02.2015).
115. Meeker W.Q., Escobar L.A. *Statistical Methods for Reliability Data*.- New York: J.Wiley and Sons,(1998).
116. O. Axelsson, *Iterative Solution Methods*, Cambridge University Press, Cambridge, 1994
117. Parzen, E. (1962). On estimation of a probability density function and mode. *Annals of Mathematical Statistics* 33, 1065–1076.
118. Pellicer-Lostao C., Morato D., Popez-Ruiz R. Modelling user's activity in a real-world complex network // *International Journal of Computer Mathematics*. Bristol: Taylor & Francis. Vol. 85, 2008. P. 1287– 1298.
119. Porshnev S.V., Koposov A.S. (2014) Quantitative estimation peculiarities of densities blur parameter of random sequences distribution in Parzen-Rosenblatt approximation. In *the World of Scientific Discoveries, Series B*, vol. 2, №1, 2014, pp. 120–129.
120. Reeves C., Rowe J. *Genetic Algorithms – principles and perspectives to GA theory*. Kluwer Academic Publishers, 2003.
121. Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Annals of Mathematical Statistics* 27, 832–837.
122. Schmitt, Lothar M (2001), *Theory of Genetic Algorithms*, *Theoretical Computer Science* 259: 1–61
123. Strateva, I., et al. 2001 Color Separation of Galaxy Types in the Sloan Digital Sky Survey Imaging Data *The Astronomical Journal* 122, 1861-1874.
124. Whitley, L. D. *Foundations of Genetic Algorithms 2*, San Mateo, California, USA: Morgan Kaufmann Publishers, 1993.
125. Wilcox, R.R. *Introduction to Robust Estimation and Hypothesis Testing*. — Academic Press, 2012.
126. Wild, C. J., and Seber, G. A. F. (2000), *Chance Encounters: A First Course in Data Analysis and Inference*, New York: Wiley, pp. 58-60.
127. Wright A. H. *Genetic algorithms for real parameter optimization*, 1991

Приложение А. Примеры ФР и ПР для различных параметров УНР

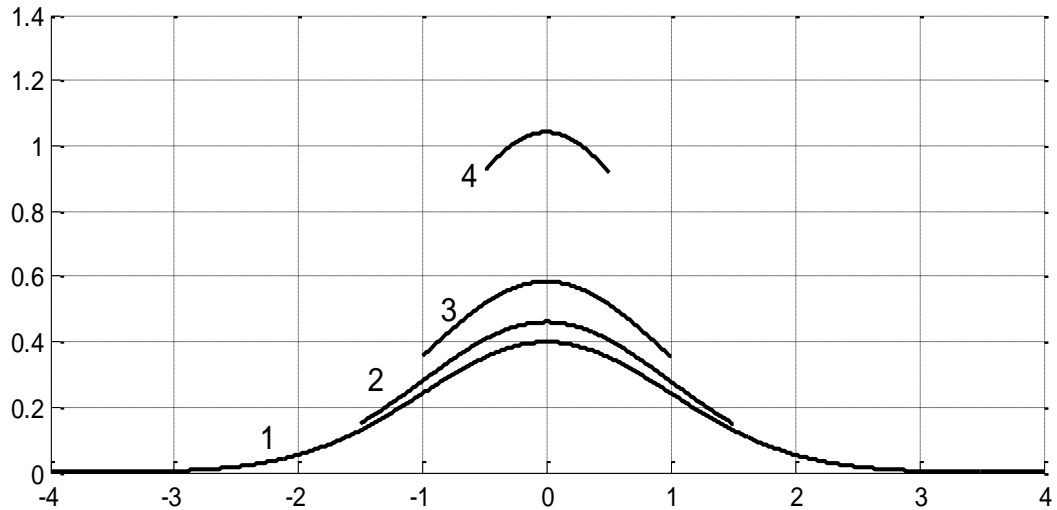


Рисунок А.1. УНР: график функции $f_{LIM}(x; x_0, \sigma, l)$:

- 1 – $x_0 = 0, \sigma = 1, 2l = 8, x \in [-l, l]$; 2 – $x_0 = 0, \sigma = 1, 2l = 3, x \in [-l, l]$;
 3 – $x_0 = 0, \sigma = 1, 2l = 2, x \in [-l, l]$; 4 – $x_0 = 0, \sigma = 1, 2l = 1, x \in [-l, l]$

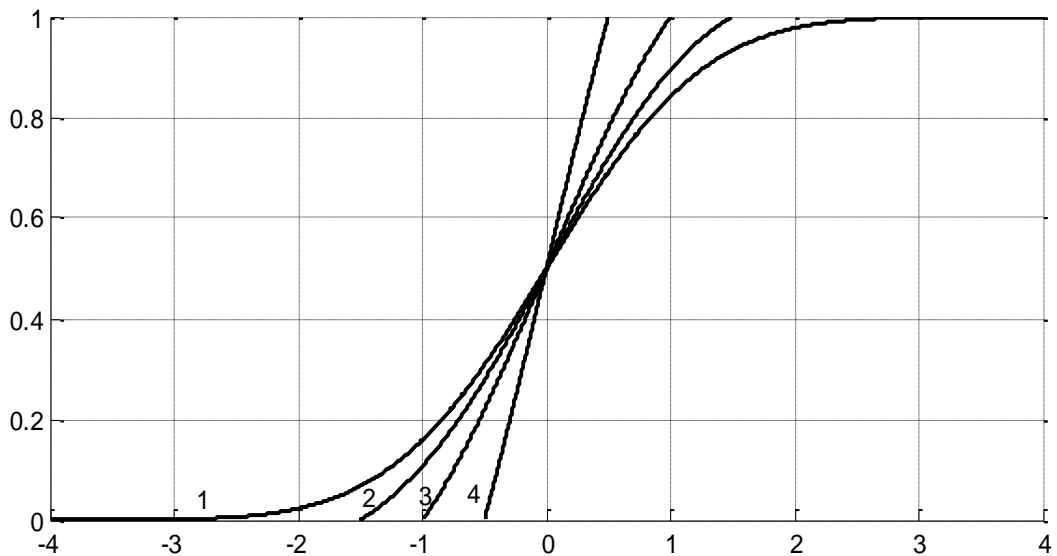


Рисунок А.2. УНР: график функции $F_{LIM}(x; x_0, \sigma, l)$:

- 1 – $x_0 = 0, \sigma = 1, 2l = 8, x \in [-l, l]$; 2 – $x_0 = 0, \sigma = 1, 2l = 3, x \in [-l, l]$;
 3 – $x_0 = 0, \sigma = 1, 2l = 2, x \in [-l, l]$; 4 – $x_0 = 0, \sigma = 1, 2l = 1, x \in [-l, l]$

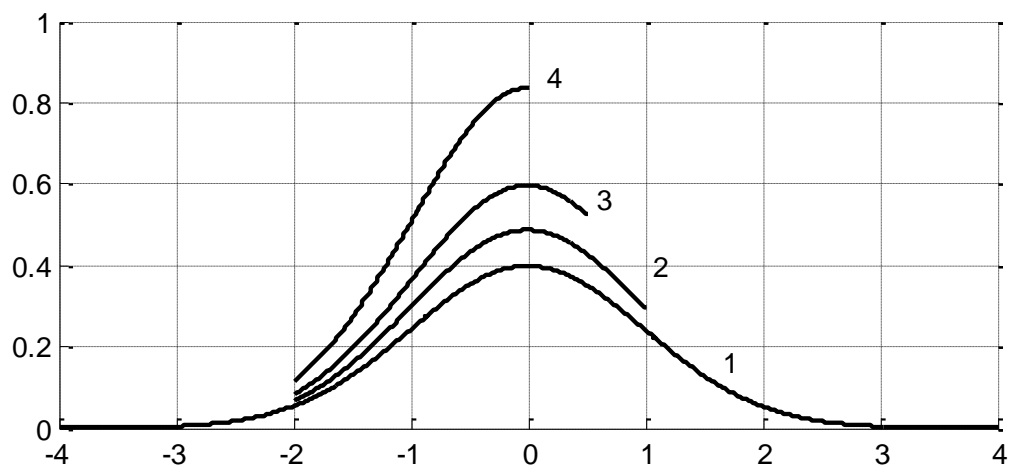


Рисунок А.3. УНР: график функции $f_{LIM}(x; x_0, \sigma, l)$:

1 – $x_0 = 0, \sigma = 1, 2l = 8, x \in [-l, l]$; 2 – $x_0 = 0, \sigma = 1, 2l = 3, x \in [-2l/3, l/3]$;
 3 – $x_0 = 0, \sigma = 1, 2l = 2.5, x \in [-8l/5, 2l/5]$; 4 – $x_0 = 0, \sigma = 1, 2l = 2, x \in [-2l, 0]$

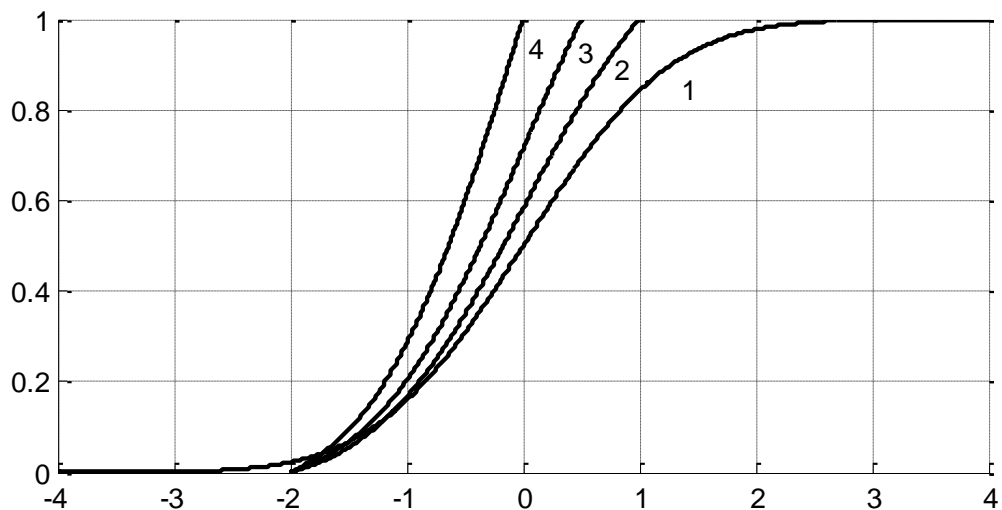


Рисунок А.4. УНР: график функции $F_{LIM}(x; x_0, \sigma, l)$:

1 – $x_0 = 0, \sigma = 1, 2l = 8, x \in [-l, l]$; 2 – $x_0 = 0, \sigma = 1, 2l = 3, x \in [-2l/3, l/3]$;
 3 – $x_0 = 0, \sigma = 1, 2l = 2.5, x \in [-8l/5, 2l/5]$; 4 – $x_0 = 0, \sigma = 1, 2l = 2, x \in [-2l, 0]$

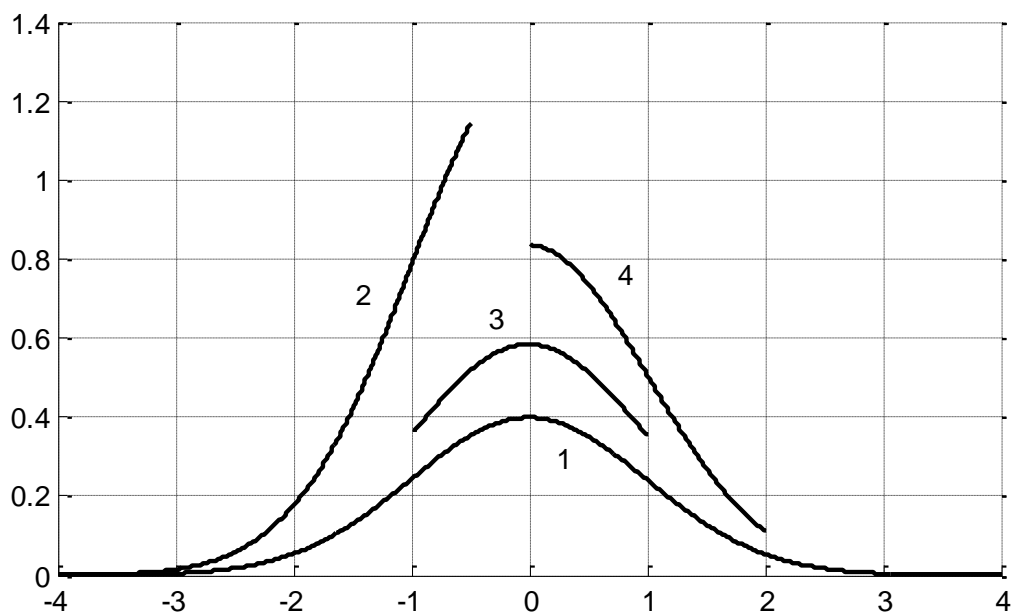


Рисунок А.5. УНР: график функции $f_{LIM}(x; x_0, \sigma, l)$:

- 1 – $x_0 = 0, \sigma = 1, 2l = 8, x \in [-l, l]$; 2 – $x_0 = 0, \sigma = 1, 2l = 4, x \in [-7l/4, -l/4,]$;
 3 – $x_0 = 0, \sigma = 1, 2l = 2, x \in [-l, l]$; 4 – $x_0 = 0, \sigma = 1, 2l = 2, x \in [0, 2l]$

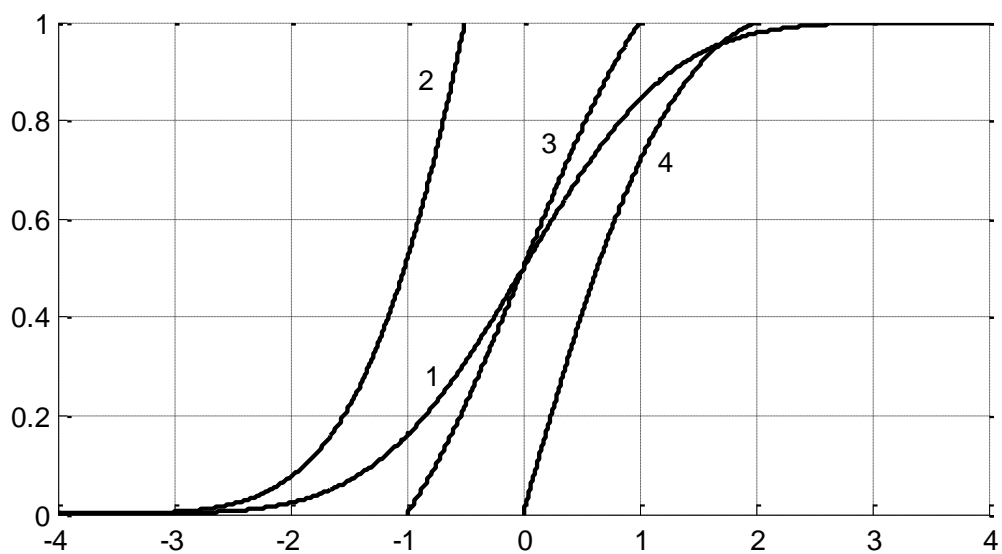


Рисунок А.6. УНР: график функции $F_{LIM}(x; x_0, \sigma, l)$:

- 1 – $x_0 = 0, \sigma = 1, 2l = 8, x \in [-l, l]$; 2 – $x_0 = 0, \sigma = 1, 2l = 4, x \in [-7l/4, -l/4,]$;
 3 – $x_0 = 0, \sigma = 1, 2l = 2, x \in [-l, l]$; 4 – $x_0 = 0, \sigma = 1, 2l = 2, x \in [0, 2l]$

Приложение Б. Результаты применения итерационного алгоритма к восстановлению ПР случайных величин с двумодальным распределением с ограниченной областью рассеяния

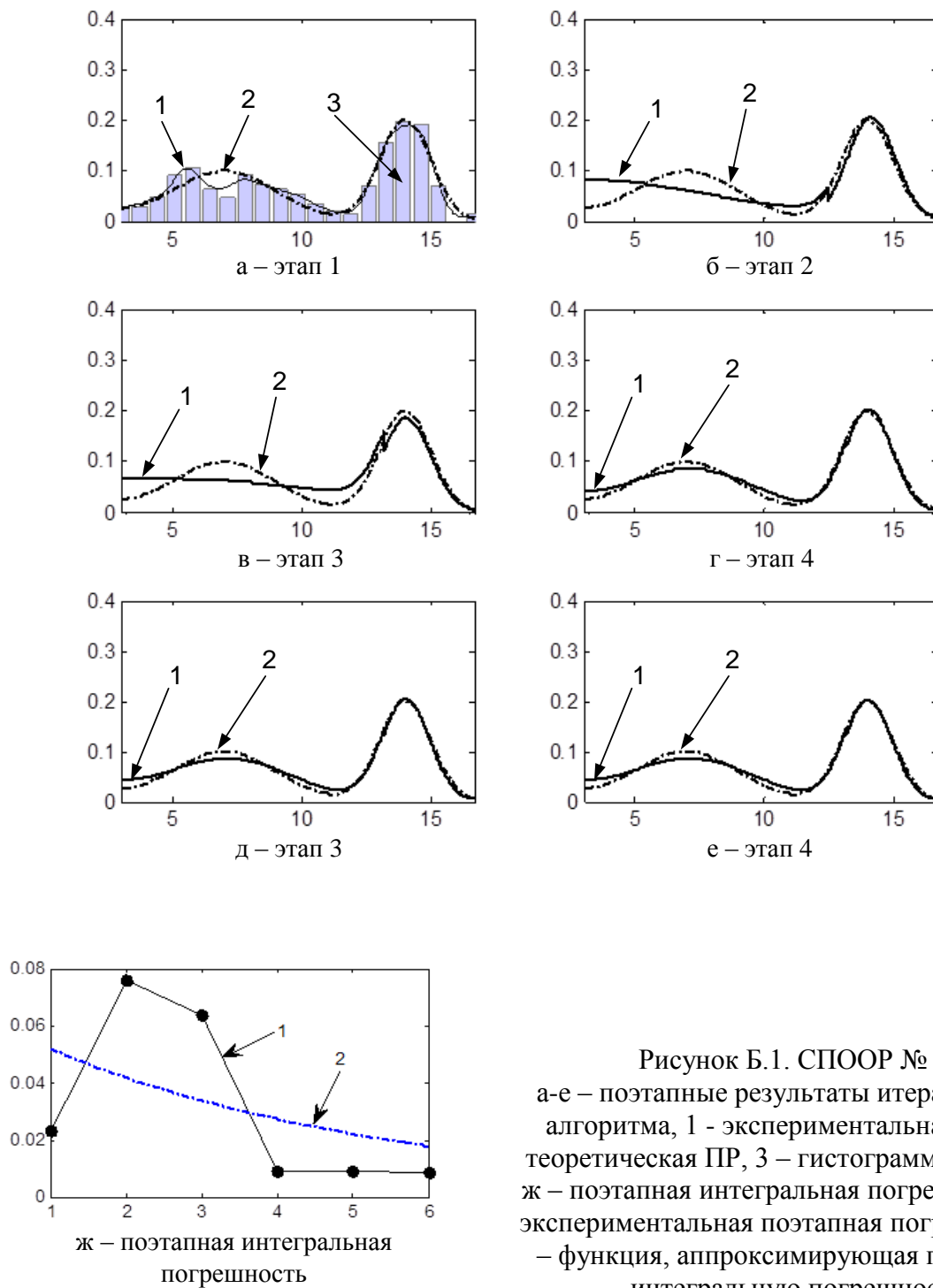


Рисунок Б.1. СПООР № 1:

а-е – поэтапные результаты итерационного алгоритма, 1 - экспериментальная ПР, 2 – теоретическая ПР, 3 – гистограмма выборки; ж – поэтапная интегральная погрешность, 1 – экспериментальная поэтапная погрешность, 2 – функция, аппроксимирующая поэтапную интегральную погрешность

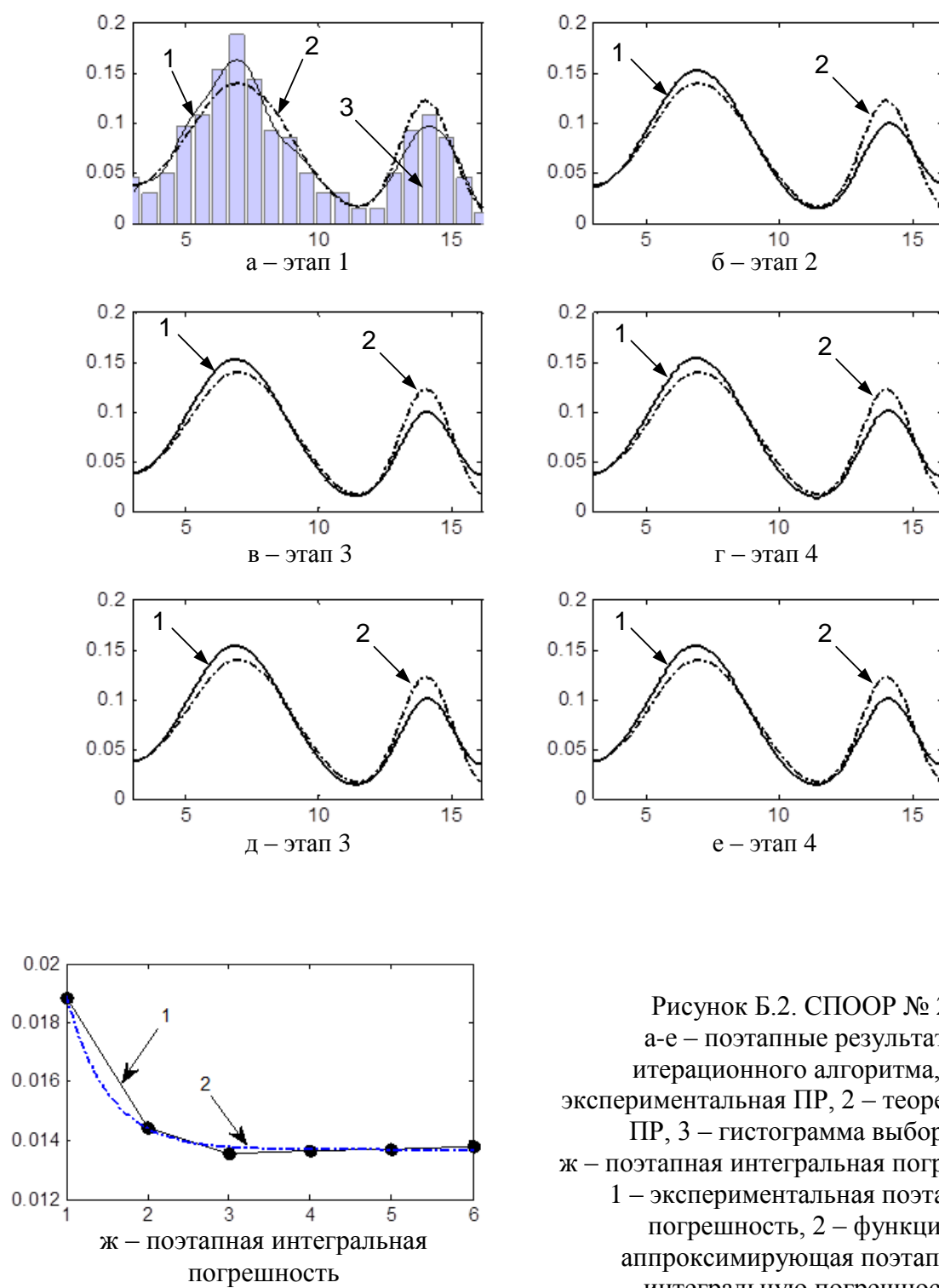
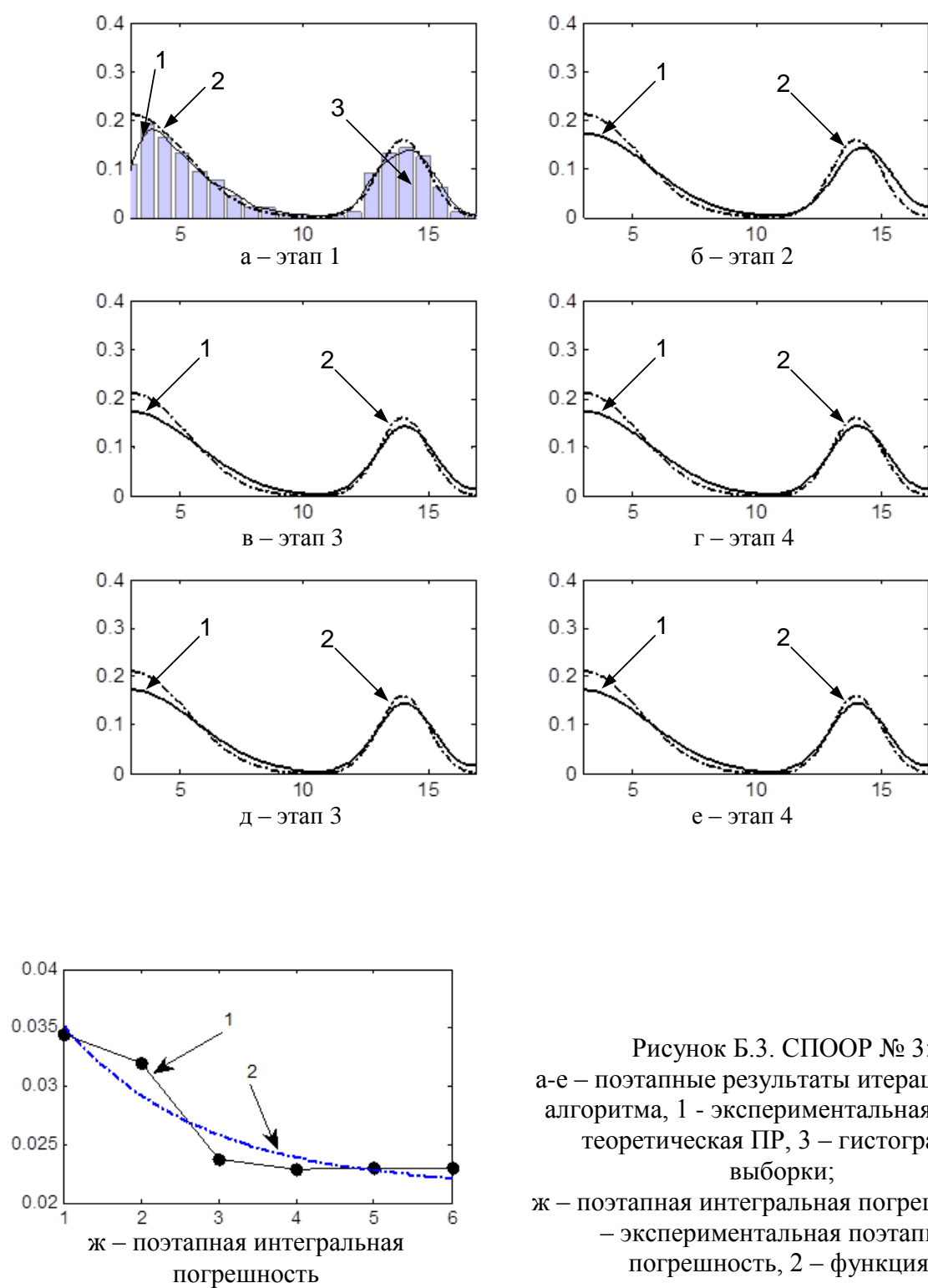


Рисунок Б.2. СПООР № 2:
а-е – поэтапные результаты
итерационного алгоритма, 1 -
экспериментальная ПР, 2 – теоретическая
ПР, 3 – гистограмма выборки;
ж – поэтапная интегральная погрешность,
1 – экспериментальная поэтапная
погрешность, 2 – функция,
аппроксимирующая поэтапную
интегральную погрешность



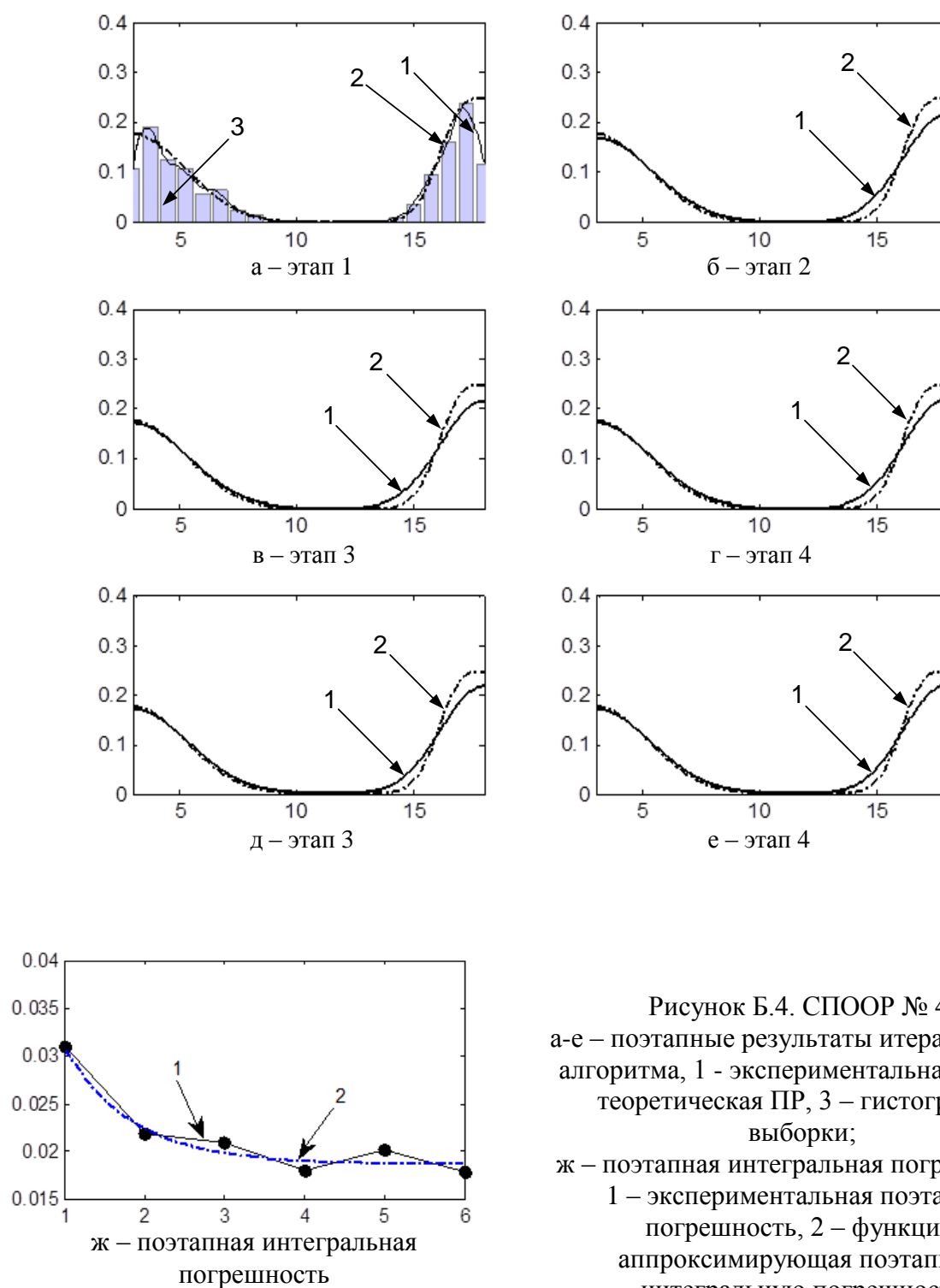


Рисунок Б.4. СПООР № 4:
а-е – поэтапные результаты итерационного алгоритма, 1 - экспериментальная ПР, 2 – теоретическая ПР, 3 – гистограмма выборки;
ж – поэтапная интегральная погрешность, 1 – экспериментальная поэтапная погрешность, 2 – функция, аппроксимирующая поэтапную интегральную погрешность

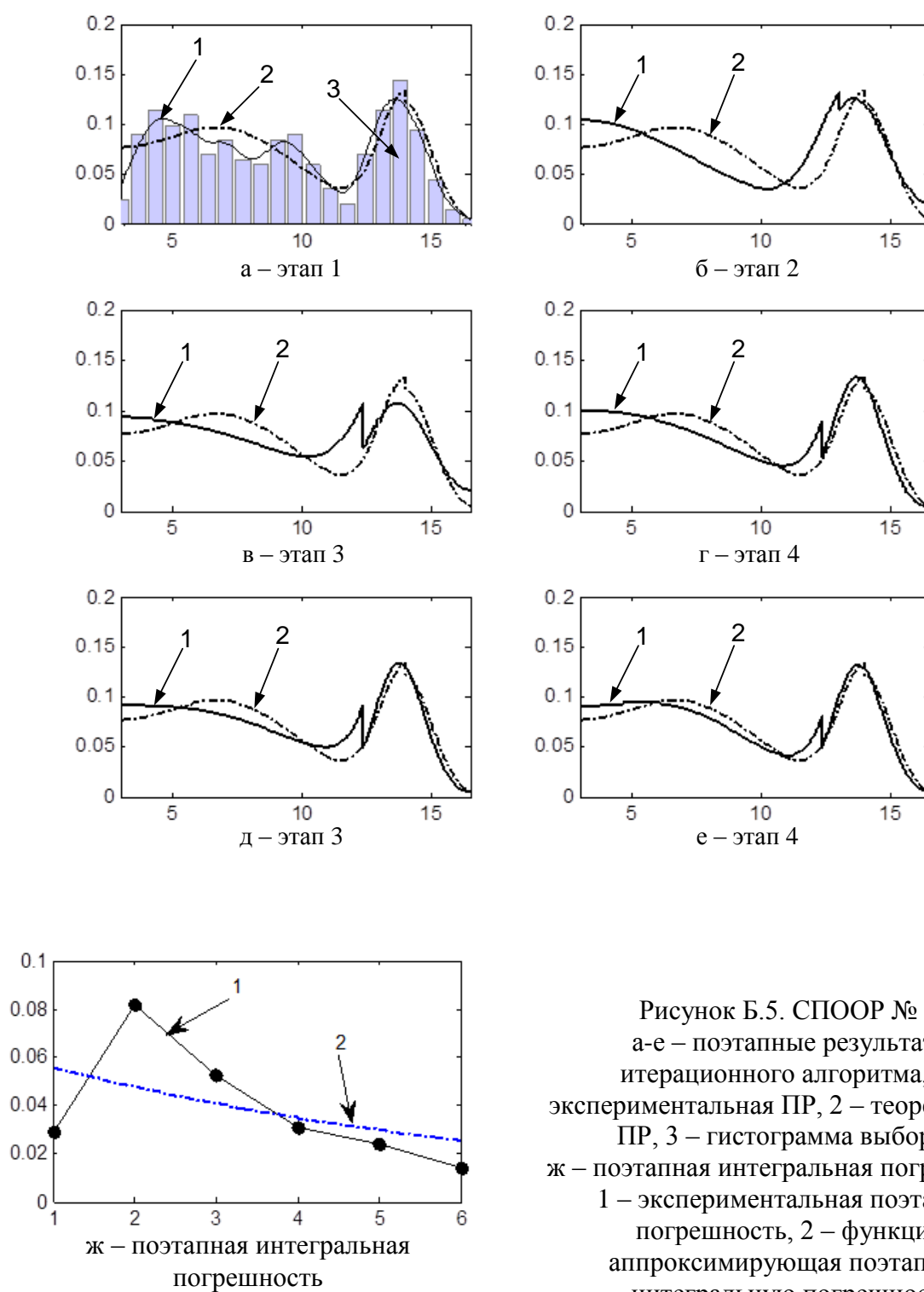


Рисунок Б.5. СПООР № 5:
а-е – поэтапные результаты
итерационного алгоритма, 1 -
экспериментальная ПР, 2 – теоретическая
ПР, 3 – гистограмма выборки;
ж – поэтапная интегральная погрешность,
1 – экспериментальная поэтапная
погрешность, 2 – функция,
аппроксимирующая поэтапную
интегральную погрешность

Приложение В. Применение метода мнимых источников и аппроксимации Розенблатта-Парзена для анализа сетевого трафика

В данной главе рассмотрен опыт применения подхода, основанного на использовании модели мнимых источников и аппроксимации Розенблатта-Парзена, для анализа особенностей потоков сетевого трафика в магистральном канале. Результаты, изложенные в данной главе, получены и изложены совместно с Д.А. Божалкиным и С.В. Поршневым.

В.1. Исследование особенностей потоков сетевого трафика в магистральном интернет-канале

Информационные потоки, передаваемые в современных компьютерных сетях (трафик), являются объектом активного изучения многих исследователей. При этом особый интерес, как с научной, так и практической точек зрения, представляет собой трафик в магистральных каналах сети Интернет, через которые проходит большое число высокоскоростных потоков пакетов.

Сегодня, несмотря на большое количество работ, посвященных изучению свойств трафика в компьютерных сетях, интерес к данной теме по-прежнему высок, что подтверждается непрекращающимся потоком публикаций по данной теме [46,32,24]. Отметим, что при анализе трафика исследователи в подавляющем большинстве работ рассматривают его как единый поток, хотя, понятно, что трафик является существенно более сложной динамической системой, состоящей из множества взаимодействующих друг с другом потоков [59]. В этой связи дальнейшее изучение свойств трафика с использованием более сложных представлений о структуре информационных потоков в компьютерных сетях является актуальной задачей.

В.1.1. Объект исследования

В качестве объекта исследования выбраны ежедневные в период с 27 октября по 2 ноября пятнадцатиминутные дампы трафика магистрального канала между США и Японией, находящиеся в хранилище данных MAWI (Measurement and Analysis on the WIDE Internet) [114]. Здесь находится как актуальная информация о текущем состоянии интернет-канала, так и архивные данные, начиная с 1999 г. Средний размер одного дампа составляет около 10 терабайт. Для сохранения конфиденциальности информации в дампе обезличены сведения о конкретных пользователях, а содержание пакетов (полезная нагрузка) заменено на сгенерированные случайным образом данные с сохранением при этом оригинального размера пакетов. Это позволяет сохранить необходимую для исследований техническую информацию (номера портов, протоколы, размер пакета и т.д.

В.1.2. Технология обработки дампа

В.1.2.1. О выборе групп пользователей, создающих информационные потоки

Следуя [118], была использована трехкомпонентная модель классификации источников трафика.

Класс А – «Слоны»: размер данных, передаваемых потоком, больше либо равен 10 Мбайт (P2P, торренты, скачивание больших файлов), время жизни таких потоков велико.

Класс В – «Мулы»: размер данных, передаваемых потоком, составляет от 0,3 до 10 Мбайт (просмотр видеороликов, прослушивание музыки, скачивание файлов небольших размеров).

Класс С – «Мыши»: размер данных, передаваемых потоком, составляет менее 0,3 Мбайт (ICQ-сообщения, просмотр WEB-страниц и т.д.).

Принадлежность источника трафика к соответствующему классу определялась на основе анализа объемов данных, передаваемых конкретным потоком. Здесь поток – совокупность данных, которая передается приложением, инициирующим поток (например, запрос клиента к серверу) как в прямом, так и в обратном направлениях.

В.1.2.2. Технология классификации потоков

Анализ дампа проводился в два этапа. На первом этапе осуществлялся семантический анализ файлов, содержащих дампы трафика, позволяющий выбрать количественные показатели трафика, удовлетворяющие заданным условиям, и передать их автоматически в рабочее пространство пакета MATLAB. Для переноса параметров пакетов дампа (необходимых, в том числе, для определения потоков выбранных классов) в окружение MATLAB (рисунок В.2 действие 3) использовался специализированный программный инструмент [55], подробно описанный в [56]. В связи с тем, что максимальный объем данных, который можно разместить в рабочем пространстве MATLAB, не нарушив при этом стабильность его работы, составляет 400 Мбайт (данный объем информации извлекается из дампа трафика, зарегистрированного на временном интервале длительностью около 2 мин), исходный файл дампа трафика делился на части соответствующей длительности, которые обрабатывались по отдельности.

После того как данные о пакетах из дампа оказывались перенесенными в MATLAB, находились потоки, которым принадлежат соответствующие пакеты. При этом принималось, что:

1. поток состоит из пакетов, передаваемых приложением, в прямом и обратном направлении;
2. суммарный объем трафика равняется суммарному объему пакетов, переданных в обоих направлениях.

Для отнесения конкретного пакета к соответствующему потоку и вычислению объема передаваемой в нем информации из дампа оказывается достаточным следующих параметров:

- время прохождения пакета через узел записи дампа;
- размер пакета (включая передаваемые данные);
- IP-адрес отправителя пакета;
- IP-адрес получателя пакета;
- порт отправителя пакета;
- порт получателя пакета;
- тип протокола.

Действительно, в соответствии с алгоритмом работы [92] сетевого (IP-адреса) и транспортного (порты) уровней модели OSI (open systems interconnection), пакеты будут принадлежать одному потоку, если они идентичны в части адреса отправителя и получателя, а также портов отправителя и получателя. При этом необходимо учитывать, что поток содержит пакеты двух типов: запросы к серверу и ответы от него (рисунок В.1).

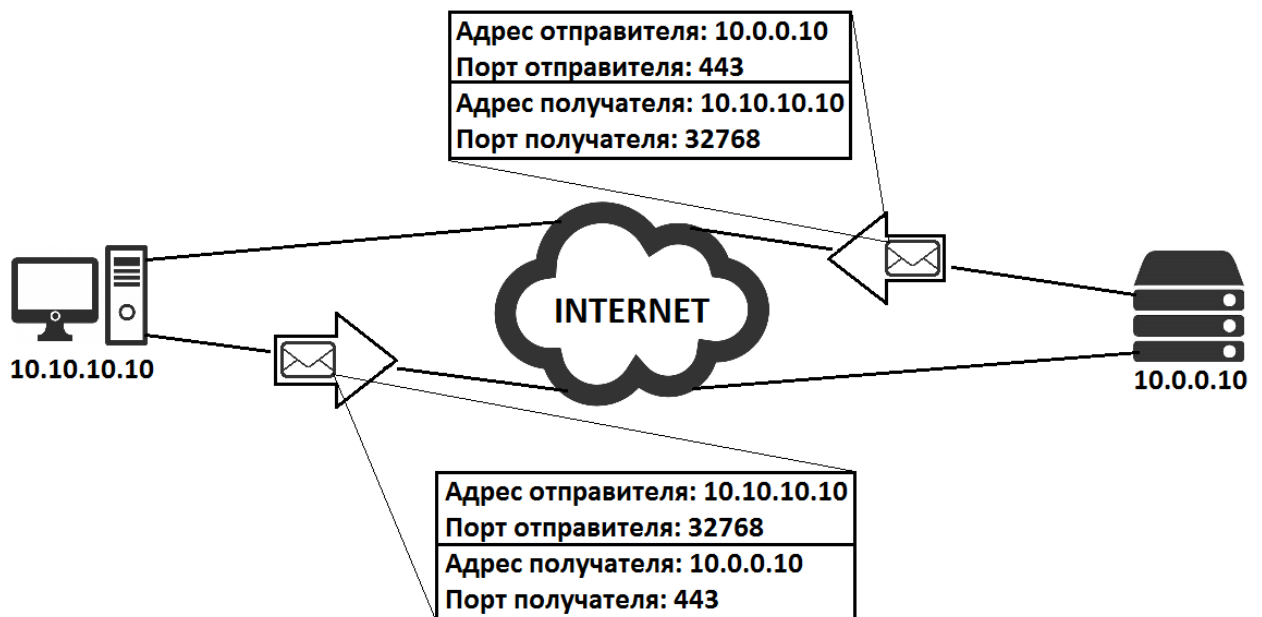


Рисунок В.1. Схема передачи пакетов одного потока

Из рисунка В.1 видно, что все пакеты, являющиеся запросами клиента (10.10.10.10) к серверу (10.0.0.10) с портом отправителя 49152, будут принадлежать одному потоку. К этому же потоку будут принадлежать и ответы сервера (10.0.0.10) клиенту (10.10.10.10) на порт 49152.

Так как клиент может иметь несколько параллельных сессий (потоков) с одним сервером, то порт получателя клиента может быть одинаковым для нескольких сессий. Порт отправителя является уникальным. Его значения назначаются динамически. Они находятся в диапазоне от 49152 до 65535. Это позволяет достоверно определять, к какому потоку принадлежит пакет. При этом порты получателя, как правило, являются общеизвестными (80-http,

53-dns) — это порты в диапазоне от 0 до 1023 или зарегистрированные от 1024 до 49151. Назначение портов контролируется администрацией адресного пространства Интернет IANA (Internet Assigned Numbers Authority) [106]. Объем данных, передаваемых потоком, вычисляется путем сложения объемов данных, передаваемых каждым пакетом этого потока (рисунок В.2 действие 9). Блок-схема последовательности действий, реализующих методику обработки дампа трафика, представлена на рисунке В.2.

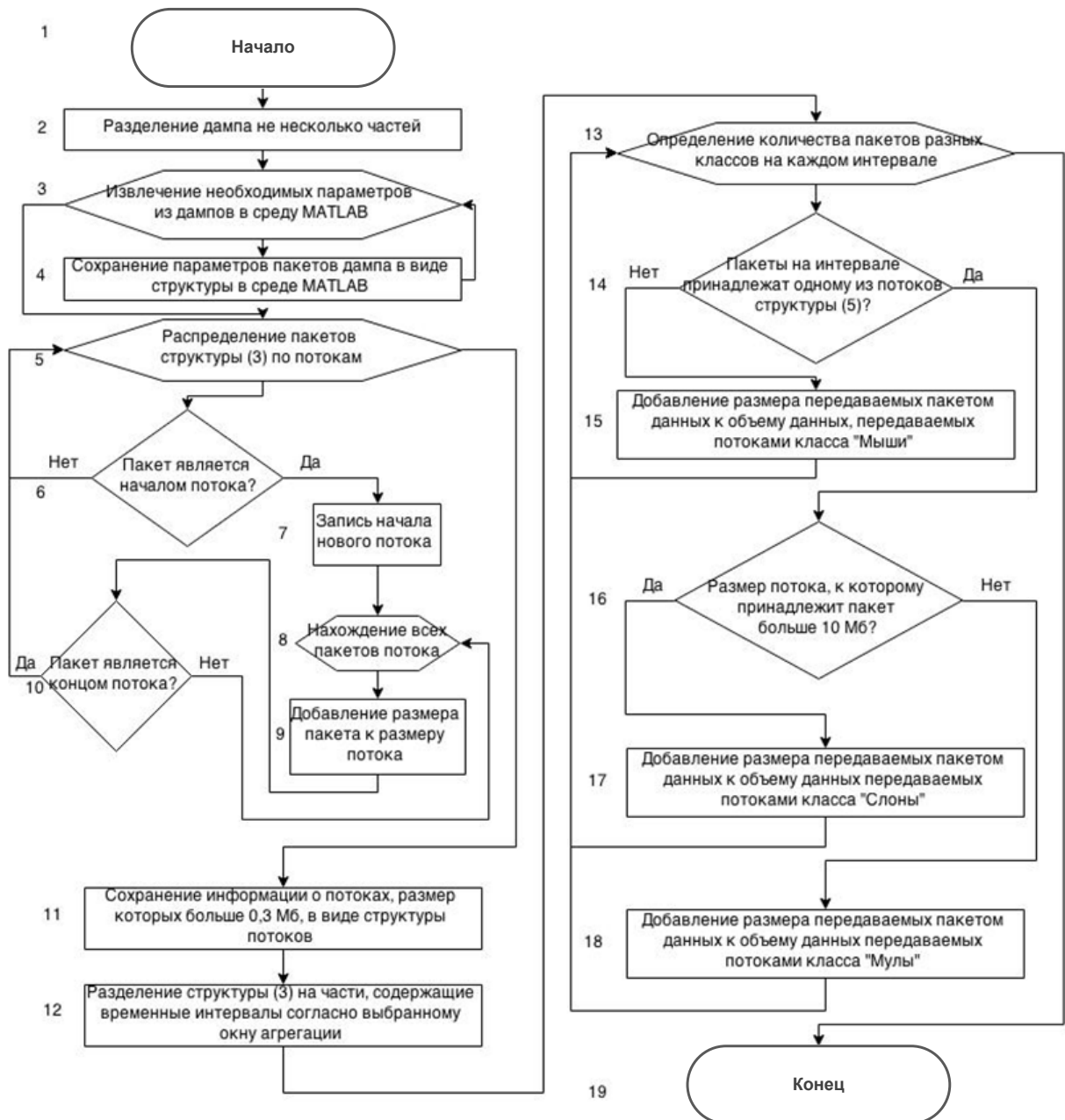


Рисунок В.2. Блок-схема основных этапов обработки дампа

В.1.2.3. Оптимизация метода классификации потоков

Для определения класса источников трафика достаточно определить потоки, относящиеся к двум классам (например, «Слонов» и «Мулов»), так как все остальные потоки будут относиться к оставшемуся классу

(«Мышей»), поэтому для ускорения процесса обработки дампа до начала поиска «Мулов» и «Слонов» из него удалялись «Мыши» – пакеты, размер потоков которых не превышает 0,3 Мбайт. Пакеты, относящиеся к данному классу, можно определить по протоколу (например, ICMP – протокол межсетевых управляющих сообщений) или порту (например, DNS – система доменных имен), что позволяет сократить время обработки дампа на треть.

В связи с тем, что большинство потоков состоит из нескольких пакетов, целесообразно определять начало и конец анализируемого потока, но не искать принадлежащие ему пакеты по всему файлу дампа, содержащего миллионы пакетов. Для определения начала и конца потока оказывается достаточным проверить не встречаются ли пакеты, относящиеся к текущему потоку, на временном интервале равном 4 с (максимальное время ожидания команды ping) до и после данного пакета. Если пакеты, удовлетворяющие данному потоку, не встречаются в течении 4 с до текущего пакета, он является началом потока, если в течении 4 с после, то он является концом потока. В остальных случаях текущий пакет будет являться частью потока, но не его началом или концом. Интервал, определяющий начало и конец потока, равный 4 с, позволяет свести к минимуму разрывы одного потока на несколько, так как взят с большим запасом (среднее время ожидания (Round-trip time) в дампе не превышает 50 мс). При определении начала и конца потока время обработки дампа в сравнении с полным прогоном уменьшается в десятки раз. Корректное определение потоков, в том случае, когда поток продолжается после перерыва более 4 с (время ожидания некоторых приложений может превышать этот интервал), можно добиться путем сравнения (алгоритм аналогичный сравнению пакетов) полученных потоков. При нахождении потоков, идентичных по параметрам (порт отправителя согласно алгоритму работы транспортного уровня является уникальным для потоков пятнадцатиминутного дампа), можно считать их одним потоком.

В.1.2.4. Аппаратные средства технологии работы с дампом

Средний размер исследуемых дампов составляет около 10ТБ, а приблизительное время обработки на одном ядре процессора составляет около 960 часов. Для сокращения времени обработки дампа в технологию его обработки была добавлена возможность проведения параллельных вычислений, что позволило использовать многоядерные процессоры и кластеры, в том числе суперкомпьютер УРАН Института математики и механики УрО РАН и, соответственно, сократить время обработки дампа в 100 раз (до 7-8 часов для одного дампа) при использовании 120 ядер центрального процессора.

При обработке были задействованы 15 вычислительных модулей со следующими параметрами:

- два 4-х ядерных процессора Intel® Xeon® E5450 (3.0 GHz);

- оперативная память 16 GB PC2-5300, Registered DDR2-667;
- кэш-память 2 x 6 MB Level 2 cache (5400 Sequence).

В.2. Экспресс-анализ статистических свойств информационных потоков в магистральном интернет-канале

В результате обработки были получены случайные последовательности (СП), содержащие значения:

- размеров пакетов L_i ;
- значения числа пакетов N_i ,
- размера потоков V_i ,

созданных в течение 15 минут в каждый из дней одной недели. К СП N_i и V_i в ходе проведенных исследований применялась процедура агрегирования, состоящая в разбиении трафика на временные интервалы длительностью $\tau = 0.1, 1.0$ с (окна агрегации) и подсчете на данном окне агрегации интегральных значений соответствующих характеристик трафика. Для краткости в обозначение каждой из СП добавлено значение окна агрегации: $N_i^{\{0.1\}}$, $N_i^{\{1.0\}}$, $V_i^{\{0.1\}}$, $V_i^{\{1.0\}}$, соответственно.

Типичные гистограммы случайных последовательностей представлены на рисунках В.3, В.4, В.5.

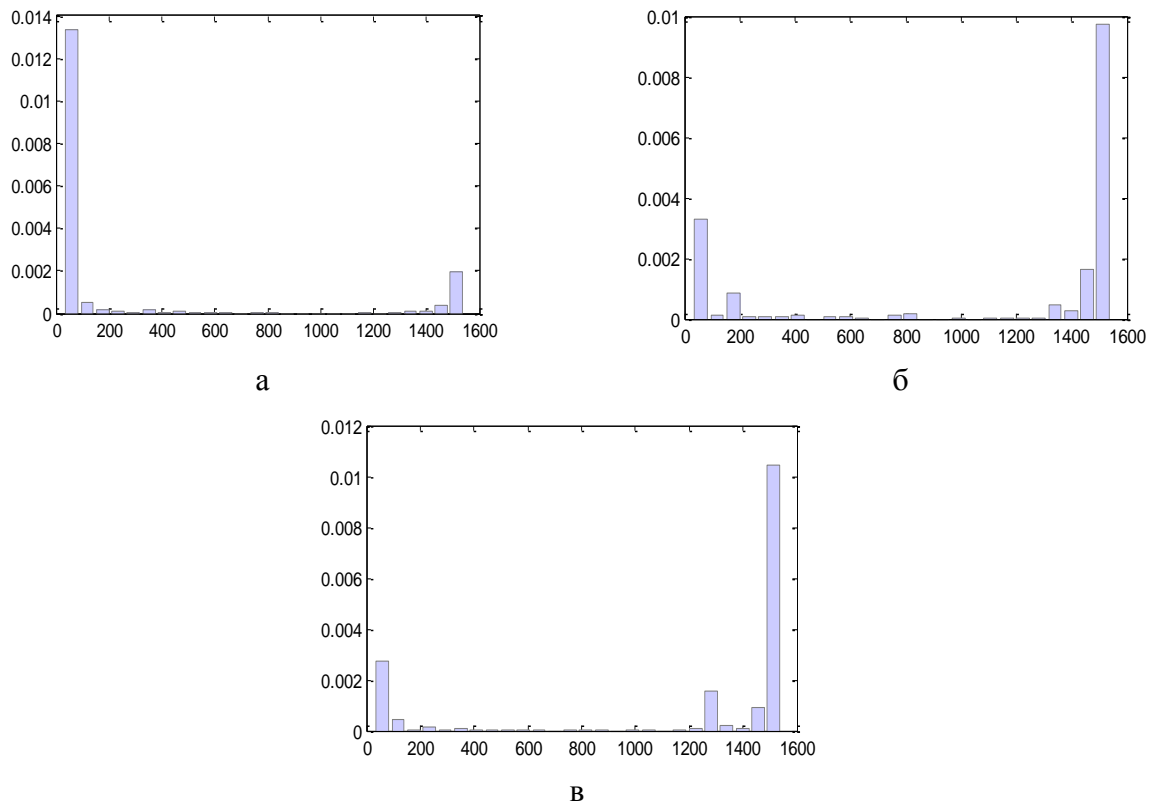


Рисунок В.3. Гистограммы СП L_i : а – «мыши», б – «мулы», в – «слоны»

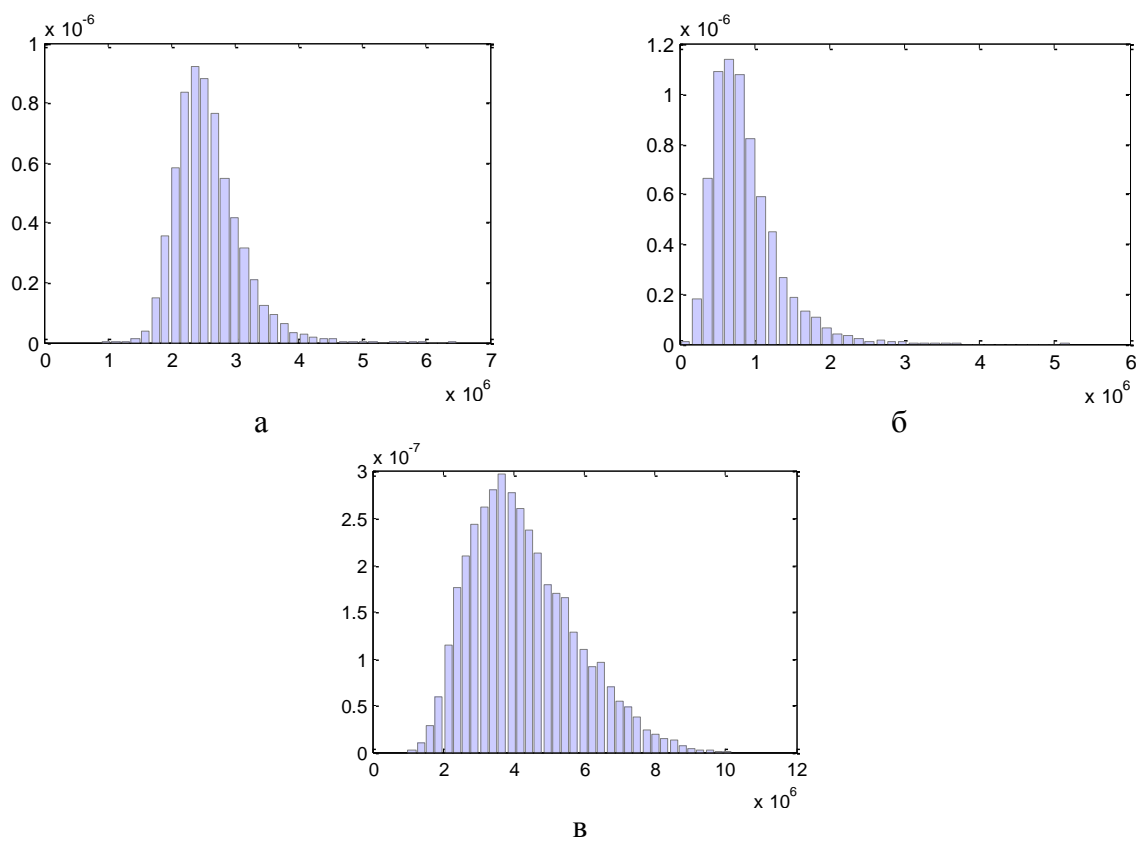


Рисунок В.4. Гистограммы СП $N_i^{\{0.1\}}$: а – «мыши», б – «мулы», в – «слоны»

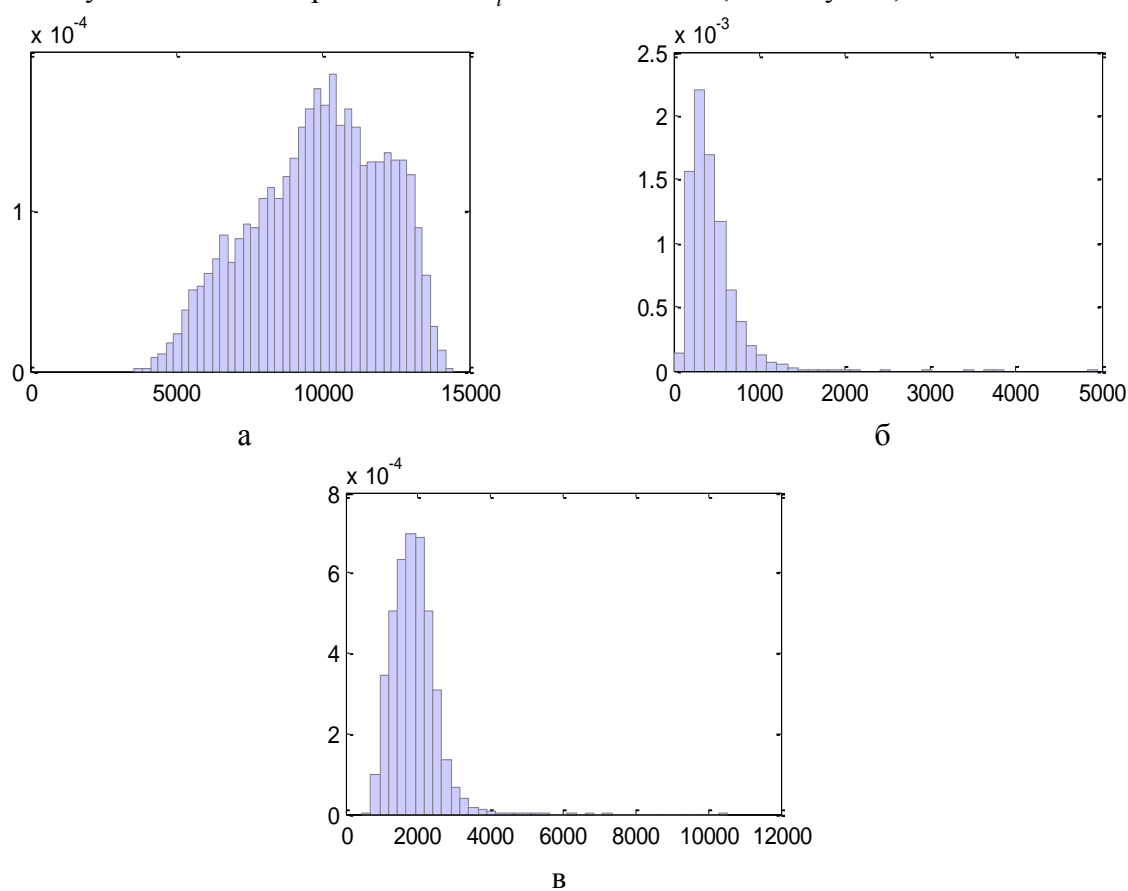


Рисунок В.5. Гистограммы СП $V_i^{\{0.1\}}$: а – «мыши», б – «мулы», в – «слоны»

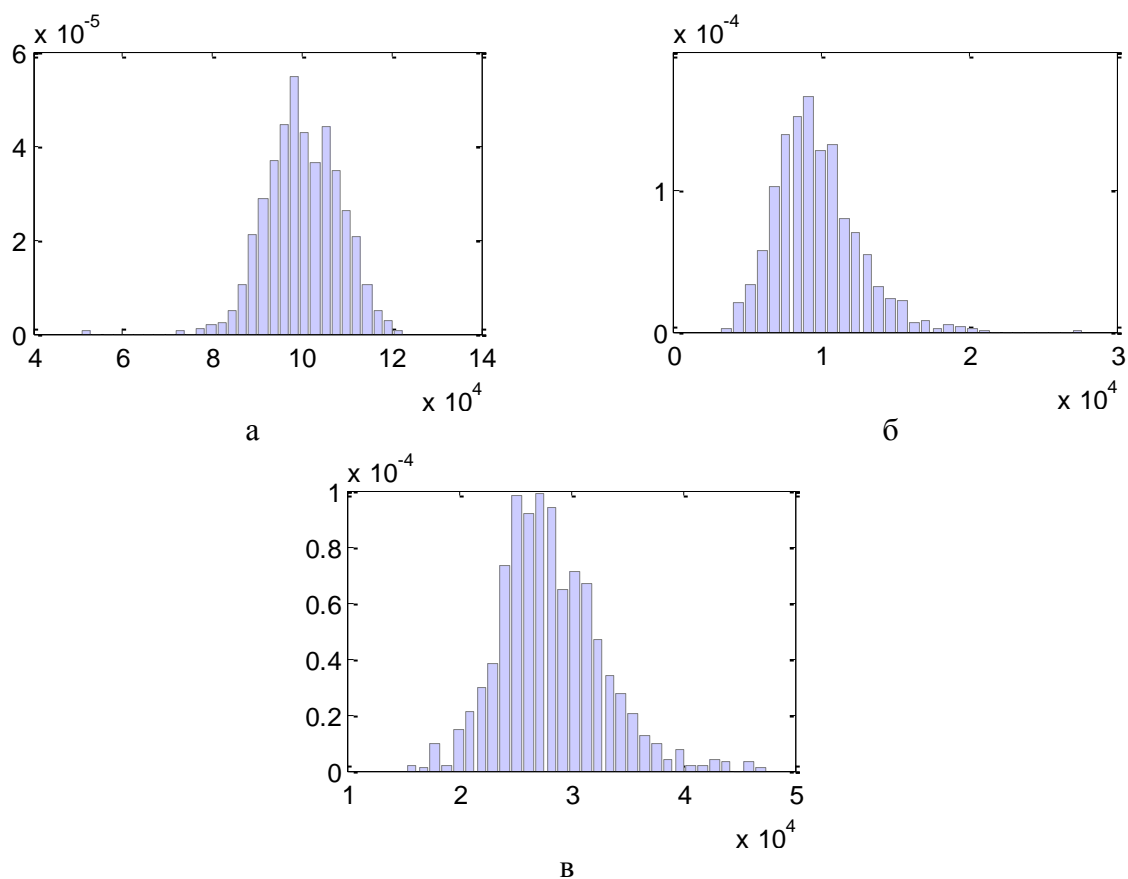


Рисунок В.6. Гистограммы СП $N_i^{1.0}$: а – «мыши», б – «мулы», в – «слоны»

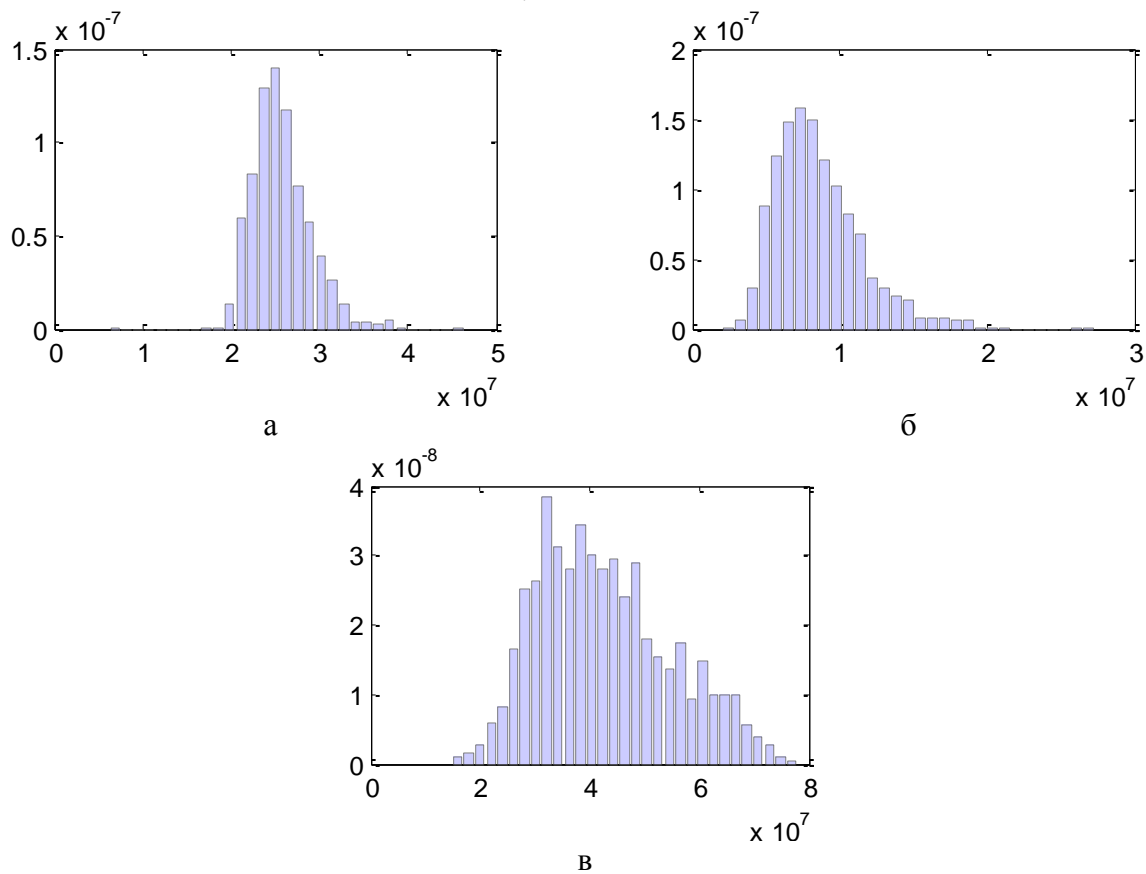


Рисунок В.7. Гистограммы СП $V_i^{1.0}$: а – «мыши», б – «мулы», в – «слоны»

Из рисунка В.3 видно, что значения СП L_i находятся в диапазоне [60; 1514] байт. Полученный результат полностью соответствует стандарту Ethernet II, включенному в 1997 г., в стандарт 802.3x [92] по которому на данный момент инкапсулируется подавляющее большинство пакетов Ethernet сетей. Действительно, в соответствии с данным стандартом кадр состоит из трех основных частей: заголовок (header) размером 14 байт (заголовок содержит: MAC-адрес получателя (Destination MAC Address); MAC-адрес отправителя (Source MAC Address); идентификатор протокола третьего уровня (IPv4, IPv6 и др.) (EtherType)); данные (data) – пакет третьего уровня размером от 46 до 1500 байт и циклический избыточный код (CRC), используемый для проверки целостности данных размером 4 байта (рисунок В.8). Таким образом, длина кадра может варьироваться от 64 до 1518 байт. (При этом кадры меньшего размера, как правило, являются результатом непредвиденных сбоев при передаче данных или ошибок в программном и аппаратном обеспечении). В соответствии с алгоритмом записи данных в архив MAWI [114], последние 4 байта (CRC) каждого пакета считаются неинформативными, а потому обрезаются сетевым оборудованием и не записываются в дампы. В этой связи максимальный и минимальный размер кадра в дампе составляют 60 и 1514 байт, соответственно.

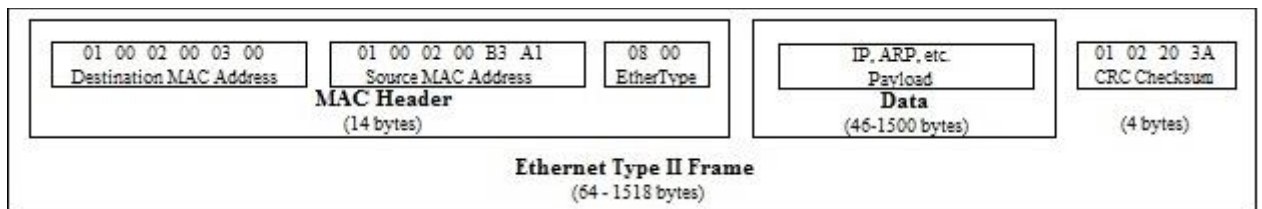


Рисунок В.8. Формат кадра Ethernet II [92]

Также необходимо отметить, что в потоках, создаваемых пользователями, относящимися к классу мыши, используются в основном пакеты небольших размеров: 60-200 байт (Рисунок В.3а), в то время как в потоках, создаваемых пользователями, относящимися к классам «мулы» и «слоны», напротив, – пакеты больших размеров (более 1000 байт). Данный результат понятен, так как небольшие пакеты используются для инициализации соединения и передачи служебной информации, а большие пакеты, имеющие максимально допустимый размер ~1500 байт, – для передачи данных.

Таким образом, полученные результаты анализа трафика в разрезе объемов созданных пакетов подтверждают работоспособность предложенной технологии работы с дампом.

Из рисунков В.4-В.7 видно:

– СП $N_i^{\{0,1\}}$, созданные «мышами», «мулами» и «слонами», имеют следующие диапазоны значений: [4800;14900]; [0;4000]; [800;10000], соответственно;

– СП $V_i^{\{0.1\}}$, созданные «мышами», «мулами» и «слонами», имеют следующие диапазоны значений: $[9 \cdot 10^5; 60 \cdot 10^5]$; $[0; 4 \cdot 10^6]$; $[1 \cdot 10^6; 10 \cdot 10^6]$, соответственно;

– СП $N_i^{\{1.0\}}$, созданные «мышами», «мулами» и «слонами», имеют следующие диапазоны значений: $[5 \cdot 10^4; 13 \cdot 10^4]$; $[0; 3 \cdot 10^4]$; $[1 \cdot 10^4; 5 \cdot 10^4]$, соответственно;

– СП $V_i^{\{1.0\}}$, созданные «мышами», «мулами» и «слонами», имеют следующие диапазоны значений: $[1 \cdot 10^7; 5 \cdot 10^7]$; $[0; 3 \cdot 10^7]$; $[1 \cdot 10^7; 8 \cdot 10^7]$, соответственно.

Таким образом, области значений СП $N_i^{\{0.1\}}$, $N_i^{\{1.0\}}$, $V_i^{\{0.1\}}$, $V_i^{\{1.0\}}$ оказываются ограниченными, что позволяет предположить, что для аппроксимации ФР и ПР данных СП следует использовать, соответственно, (1.14), (1.16).

В.3. Анализ параметров распределений случайных последовательностей N_i , V_i

Для нахождения значений параметров μ и σ в (1.14) и (1.16) случайных последовательностей N_i и V_i была использована методика, описанная в Главе 2.

В.3.1. Анализ параметров распределений количества пакетов

Результаты вычисления параметров распределений случайных последовательностей N_i представлены в таблицах В.1-В.3.

Таблица В.1. Параметры распределений СП N_i , переданных «мышами» в течение каждого из 7 дней

| i | $\tau = 0.1 \text{ с}$ | | | | $\tau = 1.0 \text{ с}$ | | | |
|-----|------------------------|---|----------------|--|------------------------|---|----------------|--|
| | $\bar{\mu}$ | $\frac{ \mu_i - \bar{\mu} }{\bar{\mu}} \cdot 100\%$ | $\bar{\sigma}$ | $\frac{ \sigma_i - \bar{\sigma} }{\bar{\sigma}} \cdot 100\%$ | $\bar{\mu}$ | $\frac{ \mu_i - \bar{\mu} }{\bar{\mu}} \cdot 100\%$ | $\bar{\sigma}$ | $\frac{ \sigma_i - \bar{\sigma} }{\bar{\sigma}} \cdot 100\%$ |
| 1 | 9813 | 1,60 | 1772 | 26,72 | 97901 | 1,33 | 8992 | 18,03 |
| 2 | | 5,92 | | 27,91 | | 6,04 | | 26,44 |
| 3 | | 2,59 | | 2,02 | | 2,36 | | 12,28 |
| 4 | | 1,30 | | 9,71 | | 1,54 | | 7,04 |
| 5 | | 0,09 | | 4,66 | | 0,25 | | 0,75 |
| 6 | | 1,07 | | 36,05 | | 0,61 | | 17,34 |
| 7 | | 9,19 | | 25,64 | | 8,98 | | 33,12 |

Таблица В.2. Параметры распределения СП N_i , переданных «мулами» в течение каждого из 7 дней

| i | $\tau = 0.1 \text{ с}$ | | | | $\tau = 1.0 \text{ с}$ | | | |
|-----|------------------------|---|----------------|--|------------------------|---|----------------|--|
| | $\bar{\mu}$ | $\frac{ \mu_i - \bar{\mu} }{\bar{\mu}} \cdot 100\%$ | $\bar{\sigma}$ | $\frac{ \sigma_i - \bar{\sigma} }{\bar{\sigma}} \cdot 100\%$ | $\bar{\mu}$ | $\frac{ \mu_i - \bar{\mu} }{\bar{\mu}} \cdot 100\%$ | $\bar{\sigma}$ | $\frac{ \sigma_i - \bar{\sigma} }{\bar{\sigma}} \cdot 100\%$ |
| 1 | 666 | 54,05 | 434 | 21,31 | 7409 | 42,41 | 2915 | 21,53 |
| 2 | | 68,70 | | 8,17 | | 51,60 | | 1,14 |
| 3 | | 45,11 | | 16,74 | | 30,23 | | 2,13 |
| 4 | | 8,58 | | 2,32 | | 1,96 | | 0,82 |
| 5 | | 53,94 | | 24,67 | | 27,68 | | 4,50 |
| 6 | | 66,41 | | 4,61 | | 47,77 | | 17,14 |
| 7 | | 56,09 | | 14,13 | | 50,75 | | 7,07 |

Таблица В.3. Параметры распределения СП N_i , переданных «слонами» в течение каждого из 7 дней

| i | $\tau = 0.1 \text{ с}$ | | | | $\tau = 1.0 \text{ с}$ | | | |
|-----|------------------------|---|----------------|--|------------------------|---|----------------|--|
| | $\bar{\mu}$ | $\frac{ \mu_i - \bar{\mu} }{\bar{\mu}} \cdot 100\%$ | $\bar{\sigma}$ | $\frac{ \sigma_i - \bar{\sigma} }{\bar{\sigma}} \cdot 100\%$ | $\bar{\mu}$ | $\frac{ \mu_i - \bar{\mu} }{\bar{\mu}} \cdot 100\%$ | $\bar{\sigma}$ | $\frac{ \sigma_i - \bar{\sigma} }{\bar{\sigma}} \cdot 100\%$ |
| 1 | 3004 | 34,42 | 722 | 10,50 | 30029 | 34,51 | 5663 | 1,45 |
| 2 | | 4,62 | | 0,25 | | 4,66 | | 0,88 |
| 3 | | 6,22 | | 17,63 | | 6,20 | | 18,46 |
| 4 | | 24,45 | | 48,54 | | 24,40 | | 54,38 |
| 5 | | 16,55 | | 14,48 | | 16,53 | | 13,35 |
| 6 | | 38,10 | | 20,30 | | 38,15 | | 13,09 |
| 7 | | 35,74 | | 35,84 | | 35,74 | | 33,85 |

Данные, представленные в таблицах В.1–В.3, позволяют оценить вариативность параметров распределений СП N_i в каждые из семи дней (Здесь $\bar{\mu}, \bar{\sigma}$ - средние по семи дням значения параметров распределения).

Из таблиц В.1–В.3 видно, что изученные СП в разные дни вне зависимости от размера окна агрегации принадлежат различным генеральным совокупностям, отличающимся друг от друга значениями параметров μ, σ . Максимальные вариации параметров значений $\mu_i, \sigma_i, i = \overline{1, 7}$ относительно значений $\bar{\mu}, \bar{\sigma}$ представлены в таблице В.4:

Таблица В.4. Максимальные вариации параметров распределений СП N_i

| Класс пользователей | $\tau = 0.1 \text{ с}$ | | $\tau = 1.0 \text{ с}$ | |
|---------------------|------------------------|--------------------------|------------------------|--------------------------|
| | $\Delta\mu_{0.1}, \%$ | $\Delta\sigma_{0.1}, \%$ | $\Delta\mu_{1.0}, \%$ | $\Delta\sigma_{1.0}, \%$ |
| «Мыши» | 9,19 | 30,05 | 8,98 | 31,20 |
| «Мулы» | 68,70 | 24,67 | 51,60 | 21,53 |
| «Слоны» | 38,10 | 48,54 | 38,15 | 54,38 |

Из таблицы В.4 видно, что вне зависимости от размера окна агрегации τ наиболее стабильными значения параметров распределения μ, σ

оказываются у класса пользователей «мыши», наименее стабильными – у класса пользователей «мулы».

Значения $\bar{\mu}_{1,0}, \bar{\sigma}_{1,0}$ и $\bar{\mu}_{0,1}, \bar{\sigma}_{0,1}$ оказались связанными между собой соотношениями $\bar{\mu}_{1,0} \approx \alpha \cdot \bar{\mu}_{0,1}$, $\bar{\sigma}_{1,0} \approx \beta \cdot \bar{\sigma}_{0,1}$, где коэффициенты α и β представлены в таблице В.5.

Таблица В.5. Значения коэффициентов α, β СП N_i

| Класс пользователей | α | β |
|---------------------|----------|---------|
| «Мыши» | 9,98 | 5,01 |
| «Мулы» | 11,12 | 6,72 |
| «Слоны» | 10,00 | 7,80 |

Из таблицы В.5 видно, что использованная процедура агрегации числа созданных пакетов не является масштабно-инвариантным преобразованием по параметру σ , а потому статистические свойства СП N_i , созданных каждым из классов пользователей, оказываются зависящими от размера окна агрегации τ . Данное обстоятельство следует принимать во внимание при статистическом моделировании информационных потоков в компьютерных сетях.

В таблицах П.Г.1-П.Г.6 представлены результаты оценки параметров распределений $\bar{\mu}_i, \bar{\sigma}_i, i = \overline{1,7}$ в каждый из дней, полученные по агрегированным СП $N_i^{\{0,1\}}, N_i^{\{1,0\}}$, созданных каждым из классов пользователей, а также относительные отклонения независимых оценок данных параметров $\mu_{i,k}, \sigma_{i,k}$, полученные по двум и четырем частям СП $N_i^{\{0,1\}}, N_i^{\{1,0\}}$, от $\bar{\mu}_i, \bar{\sigma}_i$, соответственно.

Из таблиц П.Г.1-П.Г.6 видно, что максимальные значения отклонений параметров $\Delta\mu_{\tau,k} = \max_i \left(\frac{|\mu_{\tau,i,k} - \bar{\mu}_{\tau,i}|}{\bar{\mu}_{\tau,i}} \right)$, $\Delta\sigma_{\tau,k} = \max_i \left(\frac{|\sigma_{\tau,i,k} - \bar{\sigma}_{\tau,i}|}{\bar{\sigma}_{\tau,i}} \right)$ от средних значений за день для каждого из классов пользователей оказались равными (таблица В.6):

Таблица В.6. Максимальные значения отклонений параметров распределений числа пакетов N_i от средних за день

| Класс пользователей | Окно агрегации τ | $\Delta\mu_{\tau k=\overline{1,2}}, \%$ | $\Delta\sigma_{\tau k=\overline{1,2}}, \%$ | $\Delta\mu_{\tau k=\overline{1,4}}, \%$ | $\Delta\sigma_{\tau k=\overline{1,4}}, \%$ |
|---------------------|-----------------------|---|--|---|--|
| «Мыши» | 0,1 | 4,86 | 12,63 | 10,00 | 13,61 |
| | 1,0 | 4,40 | 12,65 | 10,04 | 27,88 |
| «Мулы» | 0,1 | 78,84 | 24,00 | 109,30 | 42,20 |
| | 1,0 | 31,77 | 24,76 | 27,85 | 61,60 |
| «Слоны» | 0,1 | 11,68 | 13,93 | 19,79 | 27,19 |
| | 1,0 | 12,85 | 21,27 | 19,95 | 40,66 |

Приведенные в таблице В.6 результаты подтверждают ранее сделанный вывод о том, что наиболее вариативным, оказывается параметр μ

распределения количества «мулов» даже при его оценке по случайной последовательности, зарегистрированной в один и тот же день. При этом уменьшение количества разбиений исходной последовательности для класса пользователей «мулы» позволяет получить менее вариативную оценку параметра μ (78,84% вместо 109,30%), при практически неизменяющейся вариативности параметра σ (24,00% и 24,74%, соответственно). Для классов пользователей «мыши» и «слоны», напротив, увеличение числа разбиений исходной последовательности приводит к увеличению вариативности оценок параметра μ с 4,86% до 10,00% и с 11,68% до 19,79%, соответственно. При этом у «мышей» практически неизменной остается вариативность оценки параметра σ (12,63% и 13,61%, соответственно), а у «слонов» она изменяется в примерно в два раза (13,93% и 27,19%, соответственно).

В.3.2. Анализ параметров распределений размера потоков

Результаты вычисления параметров распределений СП V_i представлены в таблицах В.7-В.9.

Таблица В.7. Параметры распределений СП V_i , переданных «мышями» в течение каждого из 7 дней

| i | $\tau = 0.1 \text{ с}$ | | | | $\tau = 1.0 \text{ с}$ | | | |
|-----|------------------------|---|-------------------|--|------------------------|---|-------------------|--|
| | $\bar{\mu}$ | $\frac{ \mu_i - \bar{\mu} }{\bar{\mu}}, \%$ | $\bar{\sigma}$ | $\frac{ \sigma_i - \bar{\sigma} }{\bar{\sigma}}, \%$ | $\bar{\mu}$ | $\frac{ \mu_i - \bar{\mu} }{\bar{\mu}}, \%$ | $\bar{\sigma}$ | $\frac{ \sigma_i - \bar{\sigma} }{\bar{\sigma}}, \%$ |
| 1 | $2,44 \cdot 10^6$ | 2,40 | $5,56 \cdot 10^5$ | 3,30 | $2,44 \cdot 10^7$ | 2,49 | $4,08 \cdot 10^6$ | 3,76 |
| 2 | | 17,25 | | 1,05 | | 17,23 | | 6,02 |
| 3 | | 4,57 | | 10,83 | | 4,46 | | 21,23 |
| 4 | | 4,96 | | 9,89 | | 4,99 | | 17,96 |
| 5 | | 11,17 | | 19,85 | | 11,31 | | 33,50 |
| 6 | | 8,02 | | 3,52 | | 8,30 | | 15,17 |
| 7 | | 9,99 | | 32,70 | | 9,55 | | 59,77 |

Таблица В.8 Параметры распределения СП V_i , созданных «мулами» в течение каждого из 7 дней

| i | $\tau = 0.1 \text{ с}$ | | | | $\tau = 1.0 \text{ с}$ | | | |
|-----|------------------------|---|-------------------|--|------------------------|---|-------------------|--|
| | $\bar{\mu}$ | $\frac{ \mu_i - \bar{\mu} }{\bar{\mu}}, \%$ | $\bar{\sigma}$ | $\frac{ \sigma_i - \bar{\sigma} }{\bar{\sigma}}, \%$ | $\bar{\mu}$ | $\frac{ \mu_i - \bar{\mu} }{\bar{\mu}}, \%$ | $\bar{\sigma}$ | $\frac{ \sigma_i - \bar{\sigma} }{\bar{\sigma}}, \%$ |
| 1 | $7,07 \cdot 10^5$ | 51,48 | $4,68 \cdot 10^5$ | 8,26 | $7,43 \cdot 10^6$ | 45,97 | $3,39 \cdot 10^6$ | 2,25 |
| 2 | | 77,74 | | 2,63 | | 68,95 | | 1,35 |
| 3 | | 21,53 | | 7,50 | | 17,26 | | 7,68 |
| 4 | | 19,94 | | 3,56 | | 15,37 | | 9,54 |
| 5 | | 65,29 | | 28,93 | | 35,08 | | 16,44 |
| 6 | | 48,16 | | 14,24 | | 57,25 | | 8,09 |
| 7 | | 57,25 | | 14,53 | | 55,21 | | 6,41 |

Таблица В.9. Параметры распределения СП V_i , созданных «слонами» в течение каждого из 7 дней

| i | $\tau = 0.1 \text{ с}$ | | | | $\tau = 1.0 \text{ с}$ | | | |
|-----|------------------------|---|-------------------|--|------------------------|---|-------------------|--|
| | $\bar{\mu}$ | $\frac{ \mu_i - \bar{\mu} }{\bar{\mu}}, \%$ | $\bar{\sigma}$ | $\frac{ \sigma_i - \bar{\sigma} }{\bar{\sigma}}, \%$ | $\bar{\mu}$ | $\frac{ \mu_i - \bar{\mu} }{\bar{\mu}}, \%$ | $\bar{\sigma}$ | $\frac{ \sigma_i - \bar{\sigma} }{\bar{\sigma}}, \%$ |
| 1 | $3,45 \cdot 10^6$ | 43,09 | $9,32 \cdot 10^5$ | 11,61 | $3,45 \cdot 10^7$ | 43,41 | $7,41 \cdot 10^6$ | 1,16 |
| 2 | | 8,01 | | 4,31 | | 8,07 | | 5,67 |
| 3 | | 10,35 | | 15,78 | | 10,15 | | 14,87 |
| 4 | | 23,87 | | 58,04 | | 23,11 | | 68,46 |
| 5 | | 13,80 | | 13,45 | | 13,95 | | 9,83 |
| 6 | | 42,10 | | 20,67 | | 42,03 | | 13,33 |
| 7 | | 36,32 | | 42,36 | | 36,36 | | 43,26 |

Данные, представленные в таблицах В.7-В.9, позволяют оценить вариативность параметров распределений СП V_i в каждый из семи дней (здесь $\bar{\mu}, \bar{\sigma}$ - средние по семи дням значения параметров распределений СП V_i).

Из таблиц В.7-В.9 видно, что изученные СП $V_i^{\{0.1\}}, V_i^{\{1.0\}}$ вне зависимости от размера окна агрегации τ в разные дни принадлежат различным генеральным совокупностям, отличающимся друг от друга значениями параметров μ, σ . Максимальные вариации параметров значений $\mu_i, \sigma_i, i = \overline{1, 7}$ относительно значений $\bar{\mu}, \bar{\sigma}$ представлены в таблице В.10:

Таблица В.10. Максимальные вариации параметров распределений СП V_i

| Класс пользователей | $\tau = 0.1 \text{ с}$ | | $\tau = 1.0 \text{ с}$ | |
|---------------------|------------------------|--------------------------|------------------------|--------------------------|
| | $\Delta\mu_{0.1}, \%$ | $\Delta\sigma_{0.1}, \%$ | $\Delta\mu_{1.0}, \%$ | $\Delta\sigma_{1.0}, \%$ |
| «Мыши» | 17,25 | 32,70 | 17,23 | 59,77 |
| «Мулы» | 77,74 | 28,93 | 68,95 | 16,44 |
| «Слоны» | 43,09 | 58,04 | 43,41 | 68,46 |

Из таблицы В.10 видно, что вне зависимости от размера окна агрегации τ наиболее стабильными значения параметров распределения μ, σ оказываются у класса пользователей «мыши», наименее стабильными - у класса пользователей «мулы».

Значения $\bar{\mu}_{1.0}, \bar{\sigma}_{1.0}$ и $\bar{\mu}_{0.1}, \bar{\sigma}_{0.1}$ оказались связанными между собой соотношениями $\bar{\mu}_{1.0} \approx \alpha \cdot \bar{\mu}_{0.1}, \bar{\sigma}_{1.0} \approx \beta \cdot \bar{\sigma}_{0.1}$, где коэффициенты α и β представлены в таблице В.11

Таблица В.11. Значения коэффициентов α , β СП V_i

| Класс пользователей | α | β |
|---------------------|----------|---------|
| Мыши | 9,99 | 7,35 |
| Мулы | 10,52 | 7,25 |
| Слоны | 9,99 | 7,94 |

Из таблицы В.11 видно, что использованная процедура агрегации СП V_i не является масштабно-инвариантным преобразованием по параметру σ , а потому статистические свойства СП V_i , созданных каждым из классов пользователей, оказываются зависящими от размера окна агрегации τ .

В таблицах П.Д.1-П.Д.6 представлены результаты оценивания параметров распределений в каждый из дней, полученные по СП $V_i^{\{0.1\}}$, $V_i^{\{1.0\}}$, а также относительные отклонения независимых оценок параметров $\mu_{i,k}$, $\sigma_{i,k}$, полученных по двум и четырем частям исходных СП $V_i^{\{0.1\}}$, $V_i^{\{1.0\}}$.

Из таблиц П.Д.1-П.Д.6 видно, что максимальные значения отклонений параметров $\Delta\mu_{\tau,k} = \max_i \left(\frac{|\mu_{\tau,i,k} - \bar{\mu}_{\tau,i}|}{\bar{\mu}_{\tau,i}} \right)$, $\Delta\sigma_{\tau,k} = \max_i \left(\frac{|\sigma_{\tau,i,k} - \bar{\sigma}_{\tau,i}|}{\bar{\sigma}_{\tau,i}} \right)$ от средних за день для каждого из классов пользователей оказались равными (таблица В.12):

Таблица В.12. Максимальные значения отклонений параметров распределений СП V_i

| Класс пользователей | Окно агрегации τ | $\Delta\mu_{\tau k=1,2}, \%$ | $\Delta\sigma_{\tau k=1,2}, \%$ | $\Delta\mu_{\tau k=1,4}, \%$ | $\Delta\sigma_{\tau k=1,4}, \%$ |
|---------------------|-----------------------|------------------------------|---------------------------------|------------------------------|---------------------------------|
| «Мыши» | 0,1 | 9,73 | 19,52 | 55,32 | 30,69 |
| | 1,0 | 9,72 | 25,79 | 41,78 | 83,48 |
| «Мулы» | 0,1 | 49,12 | 23,47 | 69,81 | 58,15 |
| | 1,0 | 46,50 | 40,38 | 45,22 | 61,28 |
| «Слоны» | 0,1 | 19,16 | 19,87 | 28,34 | 26,70 |
| | 1,0 | 20,39 | 31,04 | 28,14 | 42,77 |

Приведенные в таблице В.12 результаты подтверждают ранее сделанный вывод о том, что наиболее вариативным, оказывается параметр μ распределения СП V_i даже при его оценке по случайной последовательности, зарегистрированной в один и тот же день.

В.3.3. Анализ параметров распределений размера пакетов L_i

В связи с тем, что вид функций плотности распределения СП L_i отличается от вида классических функций плотности распределения, для аппроксимации исследуемых плотностей распределений использовалась аппроксимация Розенблатта-Парзена. В качестве количественных показателей, характеризующих исследуемые плотности распределений, были выбраны квантили, соответствующие вероятностям: 0,25;0,5;0,75;0,95.

Результаты вычисления параметров $M_{25}, M_{50}, M_{75}, M_{95}$ распределений СП L_i , содержащих значения размеров пакетов созданных в течение пятнадцати минут каждого из дней одной недели, представлены в таблицах В.13-В.15. (Здесь \bar{M} – среднее по семи дням значения соответствующего квантиля распределения)

Таблица В.13. Параметры распределений СП L_i , созданных «мышами» в течение каждого из 7 дней

| i | \bar{M}_{25} | $\frac{ M_{25,i} - \bar{M} }{\bar{M}}, \%$ | \bar{M}_{50} | $\frac{ M_{50,i} - \bar{M} }{\bar{M}}, \%$ | \bar{M}_{75} | $\frac{ M_{75,i} - \bar{M} }{\bar{M}}, \%$ | \bar{M}_{95} | $\frac{ M_{95,i} - \bar{M} }{\bar{M}}, \%$ |
|-----|----------------|--|----------------|--|----------------|--|----------------|--|
| 1 | 81,29 | 18,14 | 108,29 | 29,15 | 196,99 | 47,77 | 1404,64 | 4,53 |
| 2 | | 19,42 | | 32,61 | | 18,66 | | 3,90 |
| 3 | | 10,99 | | 18,41 | | 35,96 | | 5,56 |
| 4 | | 17,63 | | 27,24 | | 9,07 | | 3,39 |
| 5 | | 18,14 | | 30,50 | | 54,41 | | 5,04 |
| 6 | | 24,79 | | 42,01 | | 40,07 | | 4,38 |
| 7 | | 14,56 | | 23,78 | | 70,33 | | 18,04 |

Таблица В.14. Параметры распределений СП L_i , созданных «мулами» в течение каждого из 7 дней

| i | \bar{M}_{25} | $\frac{ M_{25,i} - \bar{M} }{\bar{M}}, \%$ | \bar{M}_{50} | $\frac{ M_{50,i} - \bar{M} }{\bar{M}}, \%$ | \bar{M}_{75} | $\frac{ M_{75,i} - \bar{M} }{\bar{M}}, \%$ | \bar{M}_{95} | $\frac{ M_{95,i} - \bar{M} }{\bar{M}}, \%$ |
|-----|----------------|--|----------------|--|----------------|--|----------------|--|
| 1 | 309,36 | 40,89 | 1392,38 | 0,43 | 1462,80 | 0,07 | 1503,10 | 0,00 |
| 2 | | 6,11 | | 2,21 | | 0,87 | | 0,19 |
| 3 | | 53,11 | | 2,70 | | 0,53 | | 0,10 |
| 4 | | 23,03 | | 2,63 | | 0,97 | | 0,19 |
| 5 | | 23,03 | | 0,92 | | 0,33 | | 0,10 |
| 6 | | 124,08 | | 0,75 | | 0,07 | | 0,00 |
| 7 | | 9,87 | | 2,39 | | 1,12 | | 0,19 |

Таблица В.15. Параметры распределений СП L_i , созданных «слонами» в течение каждого из 7 дней

| i | \bar{M}_{25} | $\frac{ M_{25,i} - \bar{M} }{\bar{M}}, \%$ | \bar{M}_{50} | $\frac{ M_{50,i} - \bar{M} }{\bar{M}}, \%$ | \bar{M}_{75} | $\frac{ M_{75,i} - \bar{M} }{\bar{M}}, \%$ | \bar{M}_{95} | $\frac{ M_{95,i} - \bar{M} }{\bar{M}}, \%$ |
|-----|----------------|--|----------------|--|----------------|--|----------------|--|
| 1 | 666,01 | 93,60 | 1387,19 | 2,49 | 1459,27 | 0,81 | 1502,47 | 0,14 |
| 2 | | 80,71 | | 2,38 | | 1,11 | | 0,24 |
| 3 | | 57,48 | | 1,08 | | 0,38 | | 0,06 |
| 4 | | 35,87 | | 0,49 | | 0,11 | | 0,04 |
| 5 | | 48,75 | | 3,22 | | 1,61 | | 0,33 |
| 6 | | 51,80 | | 4,01 | | 1,58 | | 0,35 |
| 7 | | 19,59 | | 3,49 | | 1,68 | | 0,35 |

Из таблиц В.13-В.15 видно, что у изученных случайных последовательностей значения $M_{25}, M_{50}, M_{75}, M_{95}$ отличаются друг от друга.

Следовательно, данные случайные последовательности принадлежат различным генеральным совокупностям, отличающимся друг от друга. При этом максимальные вариации параметров значений $M_{25,i}, M_{50,i}, M_{75,i}, M_{95,i}, i = \overline{1,7}$ относительно значений $\bar{M}_{25}, \bar{M}_{50}, \bar{M}_{75}, \bar{M}_{95}$ составили (таблица В.16):

Таблица В.16. Максимальные значения отклонений параметров СП L_i от средних за день

| Класс пользователей | $\Delta M_{25}, \%$ | $\Delta M_{50}, \%$ | $\Delta M_{75}, \%$ | $\Delta M_{95}, \%$ |
|---------------------|---------------------|---------------------|---------------------|---------------------|
| «Мыши» | 24,79 | 42,01 | 70,33 | 18,04 |
| «Мулы» | 124,08 | 2,70 | 1,12 | 0,19 |
| «Слоны» | 93,60 | 4,01 | 1,68 | 0,35 |

Из таблицы В.16, можно сделать следующие выводы:

1. У «мышей» наиболее стабильным параметром оказывается M_{95} . Остальные параметры изменяются в диапазоне достигающим 70%.
2. У «слонов» и «мулов» наиболее стабильными параметрами оказываются M_{50}, M_{75}, M_{95} . (Диапазон их изменений не превышает 4%.) Наименее стабильным оказывается параметр M_{25} (диапазон его изменения достигает 124%).
3. Можно считать, что распределения информационных потоков в части размера пакетов, создаваемых «мулами» и «слонами» в течение одной недели, близки к стационарным распределениям.

В.4. Исследование устойчивости статистических свойств информационных потоков во времени

В результате предварительной обработки дампов были получены СП L_i, N_i, V_i , созданные каждым из выделенных классов пользователей в течение 15 минут за одни сутки с 14:00 по 14:15. Далее из СП N_i, V_i , были получены последовательности $N_i^{\{0.1\}}, V_i^{\{0.1\}}, N_i^{\{1.0\}}, V_i^{\{1.0\}}$. Учитывая, что члены данных СП упорядочены во времени, каждую из них можно рассматривать временной ряд (ВР) [3,10,28,29,90]. Количество исходных ВР для каждой из групп пользователей в каждый день равнялось 14 для СП N_i и СП V_i (7 дней \times 2 значения окна агрегации) и 7 для размера пакетов L_i (7 дней).

Результаты формального применения теста Дики-Фуллера [97] к полученным ВР для всех СП N_i, V_i и L_i показали, что изученные ВР следует считать стационарными. Однако данный результат противоречит физическим представлениям о механизмах формирования информационных потоков. Очевидно, что активность пользователей, порождающая информационные потоки, меняется случайным образом, как в течение 15-ти минутного интервала регистрации дампа трафика, так и по дням недели (например,

рабочий и выходной дни). Данный результат потребовал проведения более детального анализа статистических свойств изучаемых ВР.

Для проведения исследования стационарности ПР изучаемых ВР во времени для N_i , V_i и L_i была использована следующая методика:

1. Для каждого ВР₁ оценивались ПР в соответствии с методом мнимых источников (для N_i и V_i) и аппроксимацией Розенблатта-Парзена (для N_i , V_i и L_i).
2. Далее каждый ВР₁ делился на 2 части и для каждого из полученных ВР_{2,1}, ВР_{2,2} выполнялся п.1.
3. Каждый ВР₁ делился на 4 части и для каждого из полученных ВР_{4,1}, ВР_{4,2}, ВР_{4,3}, ВР_{4,4} выполнялся п.1.
4. Для каждой оцененной ПР, полученной в пп. 1–3 для N_i и V_i вычислялись параметры μ и σ , для L_i – квантили распределений, отвечающих доверительной вероятности 0,25, 0,50, 0,75, 0,95: $M_{25}, M_{50}, M_{75}, M_{95}$, соответственно.

В.4.1. Исследование случайной последовательности N_i на стационарность

В ходе исследования было получено оценки 294 ПР (3 класса пользователей \times 2 окна агрегации \times 7 дней \times 7 частей выборок). Примеры ПР для окон агрегации длительностями 0.1, 1.0 с представлены на рисунках В.9, В.10. **Ошибка! Источник ссылки не найден.**, соответственно.

Из рисунков В.9, В.10 видно, что ПР N_i , оцениваемые по всему ВР и его частям, оказываются отличными друг от друга, что, с нашей точки зрения, свидетельствует, вообще говоря, о его нестационарности.

Далее были получены оценки соответствия ПР анализируемых ВР найденным аппроксимациям с помощью критерия типа Колмогорова-Смирнова [33,34,47-49,73,125]. При этом нулевая гипотеза формулировалась следующим образом: найденная аппроксимация ПР соответствует функции плотности распределения исследуемой случайной последовательности. Оказалось, что для аппроксимации Розенблатта-Парзена данная гипотеза подтвердилась в 100% случаев. Полученный результат, принимая во внимание, что аппроксимация Розенблатта-Парзена относится к непараметрическим методам аппроксимации ФР и ПР СП вполне ожидаем.

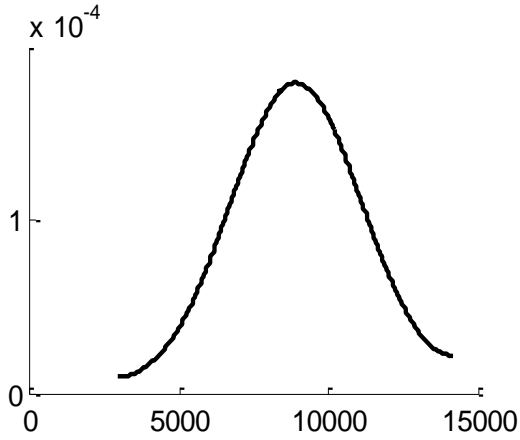
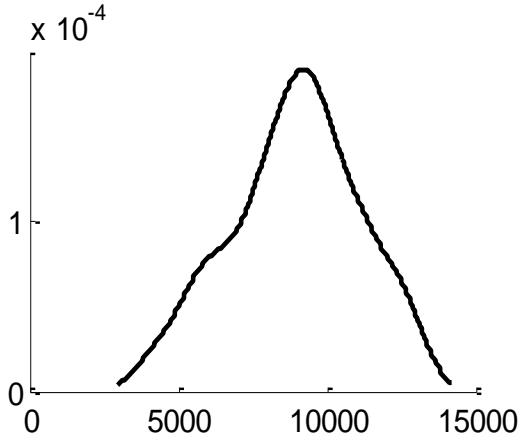
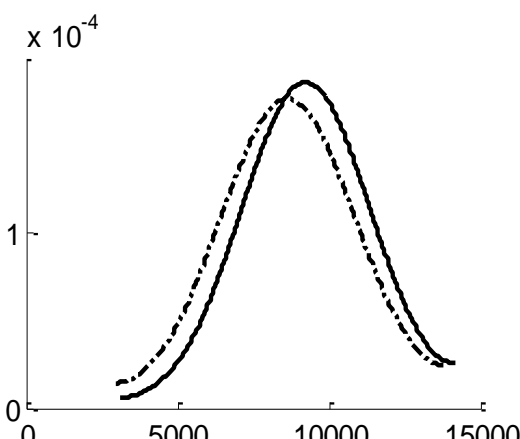
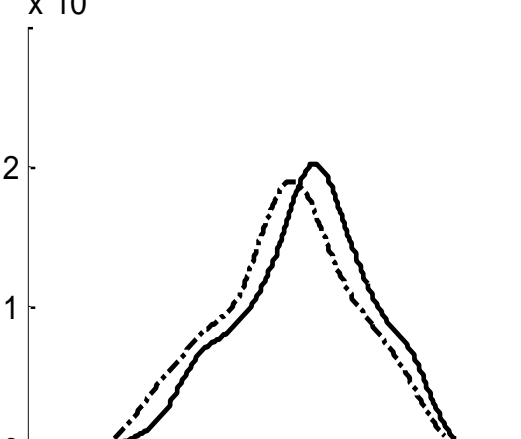
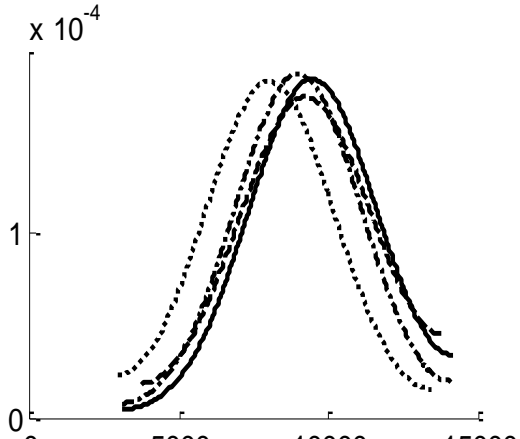
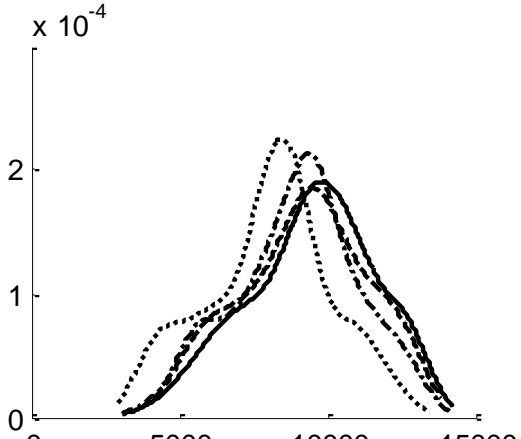
| Ряд | Метод мнимых источников | Аппроксимация Розенблатта-Парзена |
|--|---|--|
| БР ₁ |  |  |
| БР _{2,1} , БР _{2,2} |  |  |
| БР _{4,1} , БР _{4,2} , БР _{4,3} , БР _{4,4} |  |  |

Рисунок В.9. ПР $N_i^{(0,1)}$ (класс пользователей: «мыши»)

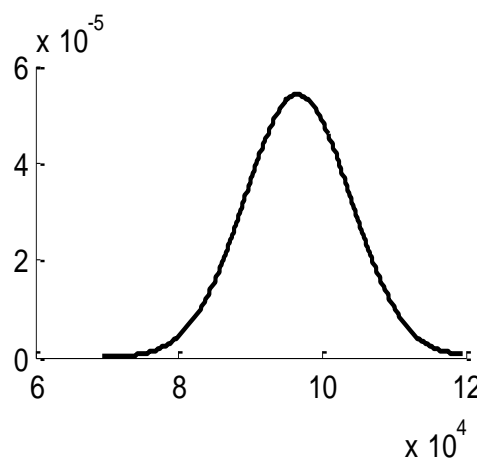
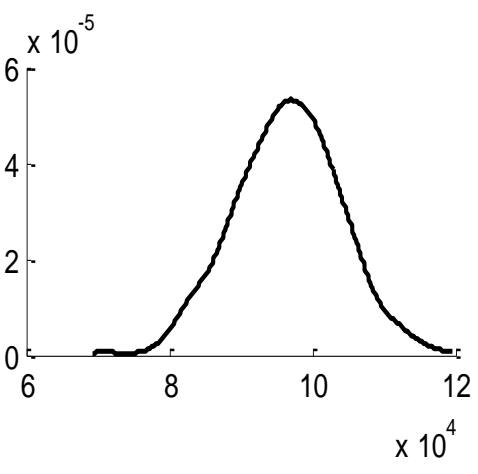
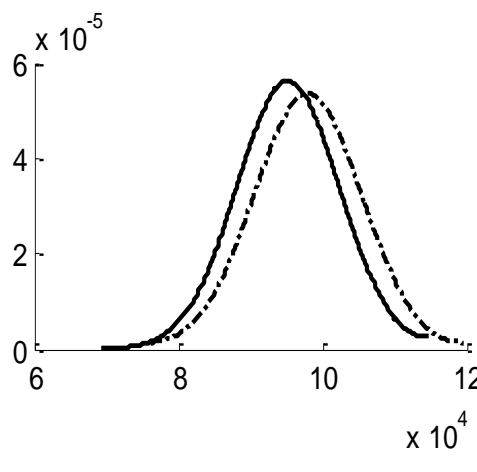
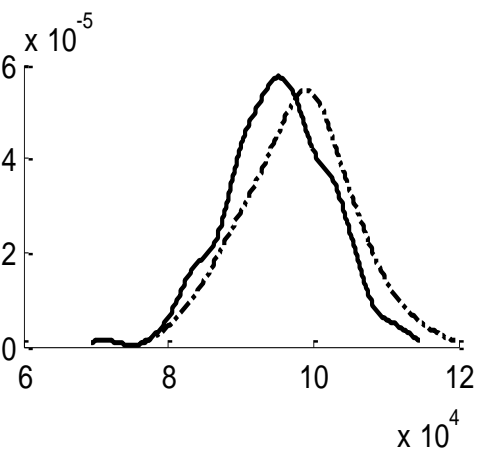
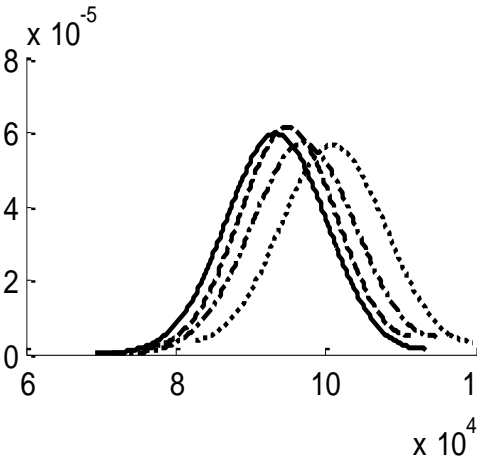
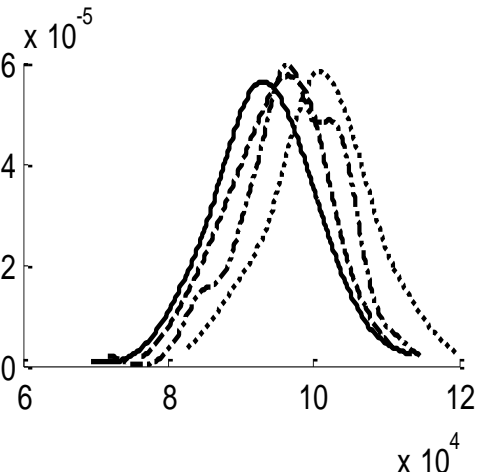
| Ряд | Метод мнимых источников | Аппроксимация Розенблатта-Парзена |
|--|---|--|
| Ряд BP ₁ |  |  |
| BP _{2,1} , BP _{2,2} |  |  |
| BP _{4,1} , BP _{4,2} , BP _{4,3} , BP _{4,4} |  |  |

Рисунок В.10. ПР $N_i^{\{1,0\}}$ (класс пользователей: «мыши»)

Обобщенные результаты оценки адекватности аппроксимации ПР изученных ВР с помощью метода мнимых источников для окон агрегации 0.1 с, 1.0 с представлены в таблицах В.17, В.18, соответственно.

Таблица В.17. Результаты применения критерия типа Колмогорова-Смирнова к оценкам ПР СП $N_i^{\{0.1\}}$

| Класс пользователей | Размер тестовой выборки | | |
|---------------------|-------------------------|-------------|--------------|
| | Полная выборка | 0,5 выборки | 0,25 выборки |
| «Мыши» | 0,0% | 21,4% | 21,4% |
| «Мулы» | 0,0% | 0,0% | 0,0% |
| «Слоны» | 0,0% | 0,0% | 17,8% |

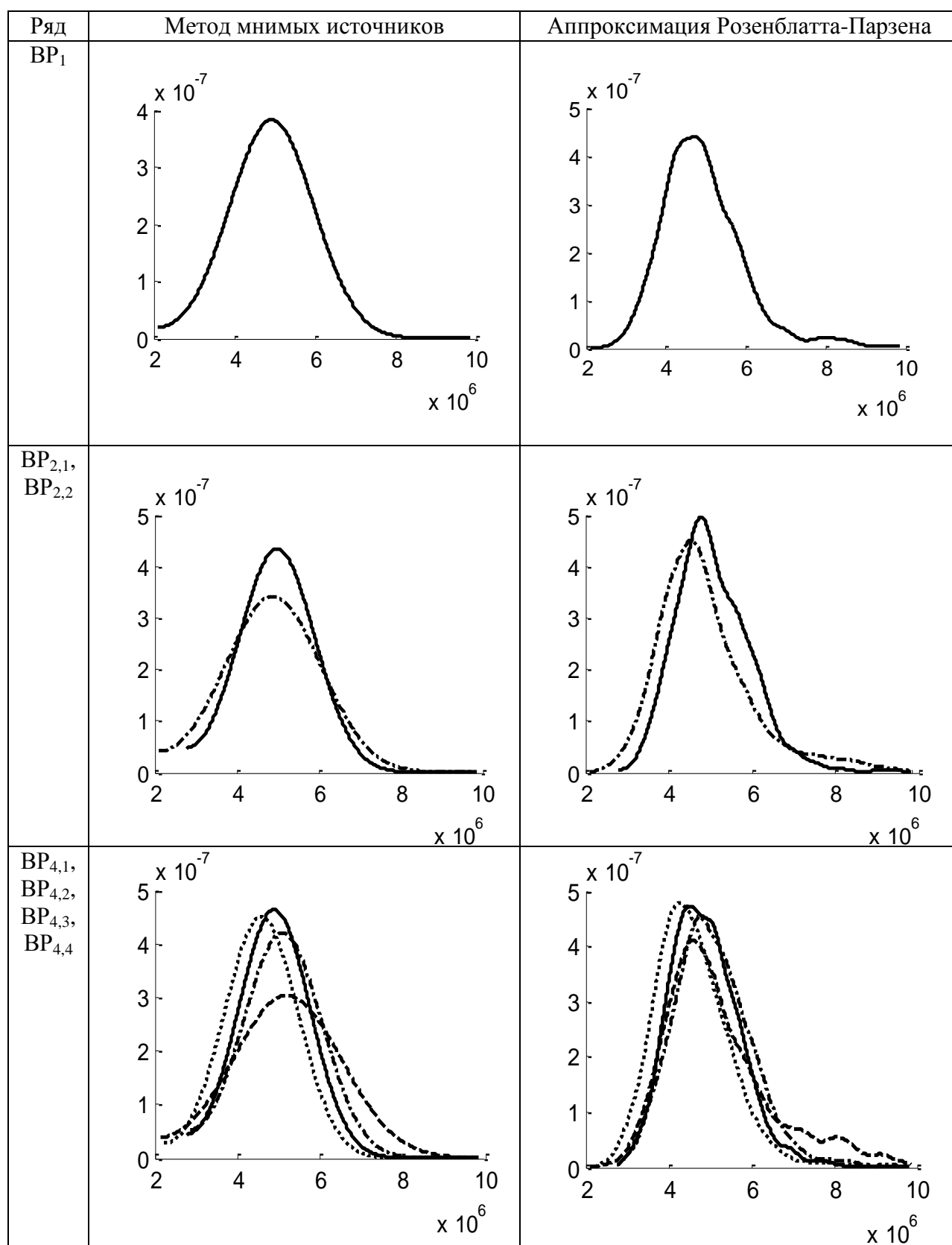
Таблица В.18. Результаты применения критерия типа Колмогорова-Смирнова к оценкам ПР СП $N_i^{\{1.0\}}$

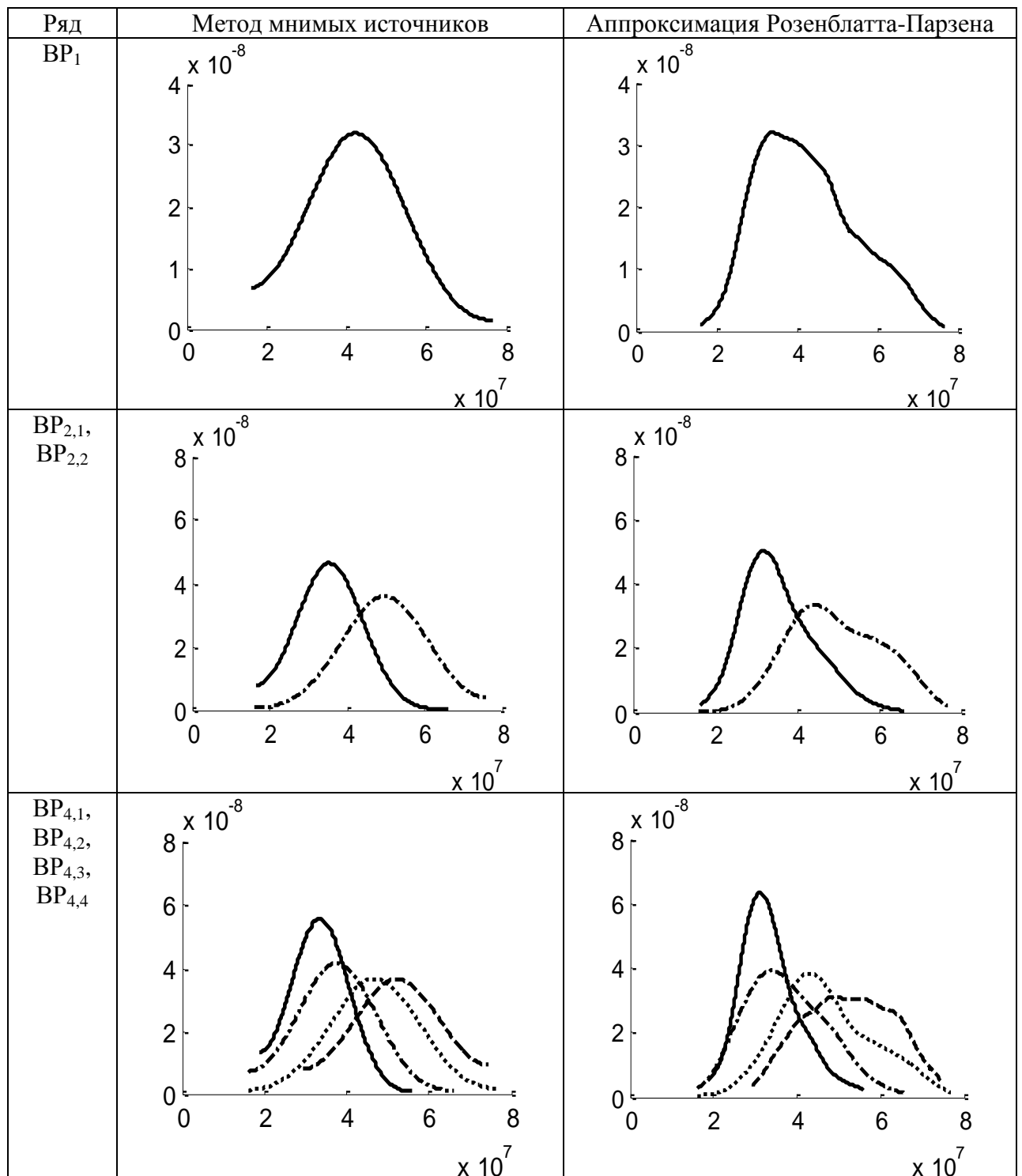
| Класс пользователей | Размер тестовой выборки | | |
|---------------------|-------------------------|-------------|--------------|
| | Полная выборка | 0,5 выборки | 0,25 выборки |
| Мыши | 100,0% | 100,0% | 100,0% |
| Мулы | 42,9% | 85,7% | 96,4 |
| Слоны | 85,7% | 85,7% | 96,4% |

Из таблиц В.17, В.18 видно, что при малых длительностях окна агрегации оценки ФР, полученные с помощью мнимых источников, оказываются неудовлетворительными. Данный результат, с нашей точки зрения, объясняется тем, что, как видно из рисунка В.9, ПР ВР, порожденных «мышами», фактически образованы суперпозицией двух независимых друг от друга распределений. Методы расщепления подобных распределений для случая, когда анализируемая выборка представляет собой смешанное распределение, рассмотрены в [54]. При больших значениях окна агрегации, оценки ПР можно считать удовлетворительными.

В.4.2. Исследование случайной последовательности V_i на стационарность

В ходе исследования было получено 294 функции плотности распределения (3 класса пользователей, 2 окна агрегации, 7 дней, 7 частей выборок). Примеры функций плотности распределения для окон агрегации 0.1, 1.0 с представлены на рисунках В.11, В.12.

Рисунок В.11. ФР СП $V_i^{(0,1)}$ (класс пользователей: «слоны»)

Рисунок В.12. ФР СП $V_i^{(1,0)}$ (класс пользователей: «слоны»)

Из рисунков В.11, В.12 видно, что ПР СП V_i , оцениваемые по всему ВР и его частям, оказываются отличными друг от друга, что с нашей точки зрения свидетельствует, вообще говоря, о его нестационарности. Для подтверждения данной гипотезы были получены оценки соответствия ПР анализируемых ВР найденным аппроксимациям. Для этого был использован критерий типа Колмогорова-Смирнова [33,34,47-49,73,125]. При этом

нулевая гипотеза формулировалась следующим образом: найденная аппроксимация ПР соответствует ПР. Оказалось, что для аппроксимации Розенблатта-Парзена данная гипотеза подтвердилась в 100% случаев. Данный результат вполне ожидаем, принимая во внимание, что аппроксимация Розенблатта-Парзена относится к непараметрическим методам аппроксимации функции плотности распределения случайных последовательностей. Обобщенные результаты оценки адекватности аппроксимации ПР изученных ВР с помощью метода мнимых источников для окон агрегации длительностью 0.1 с, 1.0 с представлены в таблицах В.19, В.20 соответственно.

Таблица В.19. Результаты применения критерия типа Колмогорова-Смирнова к оценкам ПР СП $V_i^{\{0.1\}}$

| Класс пользователей | Размер тестовой выборки | | |
|---------------------|-------------------------|-------------|--------------|
| | Полная выборка | 0,5 выборки | 0,25 выборки |
| «Мыши» | 0,0% | 0,0% | 0,0% |
| «Мулы» | 0,0% | 0,0% | 0,0% |
| «Слоны» | 0,0% | 0,0% | 28,6% |

Таблица В.20. Результаты применения критерия типа Колмогорова-Смирнова к оценкам ПР СП $V_i^{\{1.0\}}$

| Класс пользователей | Размер тестовой выборки | | |
|---------------------|-------------------------|-------------|--------------|
| | Полная выборка | 0,5 выборки | 0,25 выборки |
| Мыши | 42,9% | 71,4% | 71,4% |
| Мулы | 42,9% | 85,7% | 96,4% |
| Слоны | 28,6% | 71,4% | 92,9% |

Из таблицы В.19 видно, что при малых длительностях окна агрегации аппроксимация ПР с помощью мнимых источников оказывается неудовлетворительной. Данный результат, с нашей точки зрения, объясняется тем, что, как видно из рисунка В.12, функции плотности распределения ВР, порожденных «слонами», фактически образованы суперпозицией двух независимых друг от друга распределений. Методы расщепления подобных распределений для случая, когда анализируемая выборка представляет собой смешанное распределение, рассмотрены в [54]. При больших значениях окна агрегации, аппроксимацию функций плотности распределения изученных ВР можно считать удовлетворительной.

В.4.3. Исследование случайной последовательности L_i на стационарность

В ходе исследования были получены оценки 147 ПР распределения (3 класса пользователей \times 7 дней \times 7 частей выборок). Примеры оценок ПР представлены на рисунке В.13.

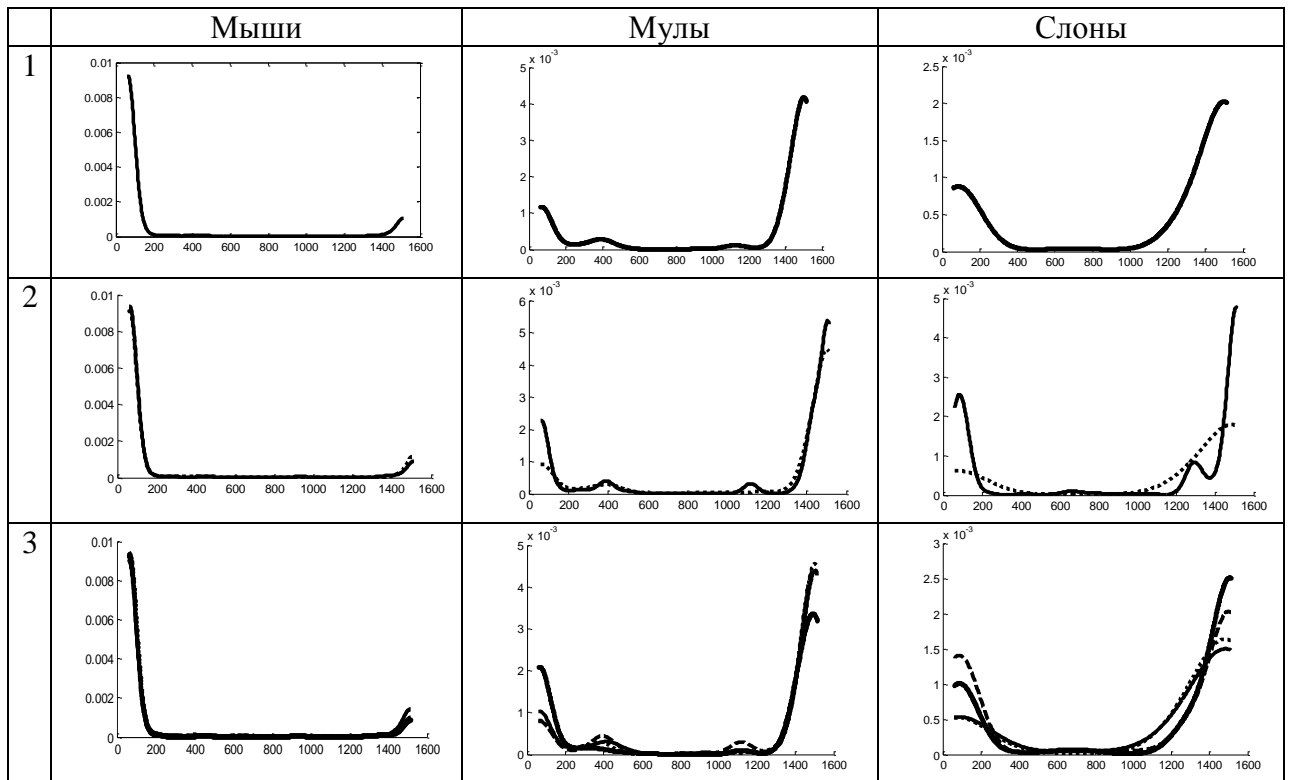


Рисунок В.13. ПР СП L_i : 1 – полная выборка; 2 – 0.5 выборки; 3 – 0.25 выборки

Результаты вычисления выбранных параметров распределений случайных последовательностей, содержащих значения размеров пакетов созданных в течение каждого из дней одной недели, представлены в таблицах П.Е.1-П.Е.3.

Из рисунка В.13 и таблиц П.Е.1-П.Е.3 видно, что ПР, оцениваемые по всей выборке и его частям, оказываются незначительно отличающимися друг от друга. Таким образом, в рамках одного дня распределение трафика, создаваемого каждым из классов пользователей, по размеру пакетов можно считать стационарным.

В.5. Выводы по главе

В.5.1. Выводы в разрезе числа пакетов

Результаты изучения статистических свойств информационных потоков в разрезе числа переданных пакетов N_i позволяют сделать следующие выводы:

1. Число пакетов, переданных каждым из классов пользователей, N_i вне зависимости от размера окна агрегации τ являются ограниченными случайными последовательностями.

2. При аппроксимации ФР и ПР распределений случайных последовательностей N_i целесообразно использовать модель СПООР.

3. Оценки параметров ФР и ПР не являются стационарными величинами, но случайным образом изменяются как по дням, так и в течение одного дня.

4. Наиболее стабильные параметры ФР и ПР, как в течение одного дня, так и в течение недели, оказываются у класса пользователей «мыши».

5. Наименее стабильные параметры распределений, как в течение одного дня, так и в течение недели, имеют пользователи классов «мулы» и «слоны», у которых отклонения параметра μ от среднего значения в течение недели могут достигать 60% и 40%, соответственно.

6. При использовании окна агрегации $\tau = 0,1$ с гипотеза о возможности аппроксимации функций плотности распределения изученных ВР с помощью метода мнимых источников в большинстве случаев отвергается критерием типа Колмогорова-Смирнова.

7. При использовании окна агрегации $\tau = 0,1$ с аппроксимации функций плотности распределения, построенные для частей ВР как с помощью метода мнимых источников, так и с помощью аппроксимации Розенблатта-Парзена, оказываются отличными, что свидетельствует о нестационарности данных ВР.

8. При использовании окна агрегации $\tau = 1$ с для аппроксимации функций плотности распределения изученных ВР методом мнимых источников обеспечивает удовлетворительное качество аппроксимации.

9. При использовании окна агрегации $\tau = 1$ с для аппроксимации функций плотности распределения изученных ВР методом мнимых источников на различных временных интервалах параметры аппроксимирующих функций оказываются отличными друг от друга, что свидетельствует о том, что даже для больших значений окон агрегации Интернет-трафик оказывается нестационарным процессом.

В.5.2. Выводы в разрезе размера потоков

Результаты изучения статистических свойств информационных потоков в разрезе размеров переданных потоков V_i следующие выводы:

1. Размер потока каждого из классов вне зависимости от размера окна агрегации τ являются ограниченными случайными последовательностями.

2. Для аппроксимации ПР и ФР изученных случайных последовательностей целесообразно модель СПООР.

3. Оценки параметров ФР и ПР не являются стационарными величинами, но могут случайным образом изменяться по дням, так и в течение одного дня, что свидетельствует о нестационарности СП V_i .

4. При использовании окна агрегации $\tau = 0,1$ с гипотеза о возможности аппроксимации ФР и ПР изученных ВР с помощью метода мнимых источников в большинстве случаев отвергается критерием Колмогорова-Смирнова.

5. Аппроксимации оценки ПР, полученные в различные дни, как с помощью метода мнимых источников, так и с помощью аппроксимации Розенблатта-Парзена, изученных ВР оказываются отличными,

6. При использовании окна агрегации $\tau = 1$ с для аппроксимации функций плотности распределения изученных ВР методом мнимых источников обеспечивает удовлетворительное качество аппроксимации.

7. При использовании окна агрегации 1 с для аппроксимации ПР СП V_i методом мнимых источников на различных временных интервалах параметры аппроксимирующих функций оказываются отличными друг от друга, поэтому даже при больших значений окон агрегации τ Интернет-трафик оказывается нестационарным процессом.

В.5.3. Выводы в разрезе размера пакетов

Результаты изучения статистических свойств информационных потоков в разрезе размера пакетов L_i , позволяют сделать следующие выводы:

1. В информационных потоках, создаваемых, «мышами», преобладающими являются пакеты размером не более 200 байт.

2. В информационных потоках, создаваемых «мулами» и «слонами», преобладающими являются пакеты размером ~ 1500 байт.

3. Информационные потоки, создаваемые каждым из классов пользователей, на пятнадцатиминутных интервалах имеют стационарные во времени распределения по размеру передаваемых пакетов.

4. Наиболее стабильные во времени параметры ФР и ПР информационных потоков по размеру пакетов в течение недели оказываются у «слонов» и «мулов», наименее стабильные – у «мышей».

В.5.4. Общие выводы

1. Предложенные в разделе 2 и разделе 3 методики оценки параметров распределений позволили провести анализ свойств сетевого трафика на примере магистрального интернет-канала между США и Японией.

2. Проведены исследования статистических свойств информационных потоков в Интернет-канале в разрезе размера пакетов, количества переданных пакетов и размера потоков.

3. Показана целесообразность применения модели СПООР и разработанных в диссертации методик, для анализа зависимостей количества пакетов и размера потоков трафика от времени.

4. Показана целесообразность использования аппроксимации Розенблатта-Парзена для анализа зависимостей размера пакетов трафика от времени.

Результаты исследования, приведенного в данной главе, опубликованы в [57, 58]

**Приложение Г. Результаты оценки параметров
распределений по каждому из дней, полученные по
полной выборке агрегированных последовательностей в
разреze количества пакетов**

Таблица Г.1. Параметры распределений СП $N_i^{\{0.1\}}$, переданных
«мышами» в каждый из дней

| i | $\bar{\mu}_i$ | k | $\frac{ \mu_{i,k} - \bar{\mu}_i }{\bar{\mu}_i}, \%$ | k | $\frac{ \mu_{i,k} - \bar{\mu}_i }{\bar{\mu}_i}, \%$ | $\bar{\sigma}_i$ | k | $\frac{ \sigma_{i,k} - \bar{\sigma}_i }{\bar{\sigma}_i}, \%$ | k | $\frac{ \sigma_{i,k} - \bar{\sigma}_i }{\bar{\sigma}_i}, \%$ |
|-----|---------------|-----|---|-----|---|------------------|-----|--|-----|--|
| 1 | 9656 | 1 | 1,40 | 1 | 3,28 | 1299 | 1 | 0,22 | 1 | 4,37 |
| | | | | 2 | 0,45 | | | | 2 | 5,38 |
| | | 2 | 1,48 | 3 | 1,69 | | 2 | 0,29 | 3 | 1,06 |
| | | | | 4 | 4,69 | | | | 4 | 3,14 |
| 2 | 10394 | 1 | 0,06 | 1 | 0,01 | 1278 | 1 | 6,76 | 1 | 7,16 |
| | | | | 2 | 0,06 | | | | 2 | 7,00 |
| | | 2 | 0,08 | 3 | 0,96 | | 2 | 6,94 | 3 | 7,12 |
| | | | | 4 | 0,74 | | | | 4 | 8,62 |
| 3 | 10067 | 1 | 0,67 | 1 | 0,40 | 1736 | 1 | 6,49 | 1 | 3,12 |
| | | | | 2 | 1,67 | | | | 2 | 11,23 |
| | | 2 | 0,83 | 3 | 0,82 | | 2 | 1,66 | 3 | 2,67 |
| | | | | 4 | 0,39 | | | | 4 | 4,54 |
| 4 | 9940 | 1 | 4,86 | 1 | 8,55 | 1600 | 1 | 6,83 | 1 | 6,10 |
| | | | | 2 | 0,93 | | | | 2 | 3,37 |
| | | 2 | 4,06 | 3 | 3,66 | | 2 | 5,68 | 3 | 13,61 |
| | | | | 4 | 4,39 | | | | 4 | 4,33 |
| 5 | 9804 | 1 | 0,08 | 1 | 0,84 | 1855 | 1 | 12,63 | 1 | 6,90 |
| | | | | 2 | 0,12 | | | | 2 | 11,57 |
| | | 2 | 0,74 | 3 | 0,28 | | 2 | 0,34 | 3 | 3,47 |
| | | | | 4 | 2,17 | | | | 4 | 9,92 |
| 6 | 9918 | 1 | 0,15 | 1 | 0,50 | 2411 | 1 | 0,67 | 1 | 0,33 |
| | | | | 2 | 0,85 | | | | 2 | 1,37 |
| | | 2 | 0,03 | 3 | 0,14 | | 2 | 1,38 | 3 | 0,81 |
| | | | | 4 | 0,34 | | | | 4 | 0,45 |
| 7 | 8911 | 1 | 3,42 | 1 | 6,22 | 2227 | 1 | 3,70 | 1 | 2,97 |
| | | | | 2 | 0,69 | | | | 2 | 4,43 |
| | | 2 | 3,31 | 3 | 3,56 | | 2 | 1,51 | 3 | 2,43 |
| | | | | 4 | 10,00 | | | | 4 | 2,31 |

Таблица Г.2. Параметры распределений СП $N_i^{\{0,1\}}$, переданных «мулами» в каждый из дней

| i | $\bar{\mu}_i$ | k | $\frac{ \mu_{i,k} - \bar{\mu}_i }{\bar{\mu}_i}, \%$ | k | $\frac{ \mu_{i,k} - \bar{\mu}_i }{\bar{\mu}_i}, \%$ | $\bar{\sigma}_i$ | k | $\frac{ \sigma_{i,k} - \bar{\sigma}_i }{\bar{\sigma}_i}, \%$ | k | $\frac{ \sigma_{i,k} - \bar{\sigma}_i }{\bar{\sigma}_i}, \%$ |
|-----|---------------|-----|---|-----|---|------------------|-----|--|-----|--|
| 1 | 1026 | 1 | 1,95 | 1 | 5,28 | 527 | 1 | 4,29 | 1 | 6,33 |
| | | | | 2 | 8,62 | | | | 2 | 27,30 |
| | | 2 | 1,59 | 3 | 11,28 | | 2 | 10,80 | 3 | 37,42 |
| | | | | 4 | 7,50 | | | | 4 | 27,81 |
| 2 | 1124 | 1 | 7,54 | 1 | 11,04 | 399 | 1 | 1,99 | 1 | 2,34 |
| | | | | 2 | 4,41 | | | | 2 | 1,17 |
| | | 2 | 8,39 | 3 | 1,13 | | 2 | 4,26 | 3 | 3,43 |
| | | | | 4 | 15,13 | | | | 4 | 10,00 |
| 3 | 967 | 1 | 2,79 | 1 | 1,39 | 362 | 1 | 0,69 | 1 | 15,07 |
| | | | | 2 | 6,47 | | | | 2 | 10,70 |
| | | 2 | 2,45 | 3 | 0,19 | | 2 | 0,76 | 3 | 3,39 |
| | | | | 4 | 4,12 | | | | 4 | 1,81 |
| 4 | 724 | 1 | 5,64 | 1 | 6,01 | 424 | 1 | 6,31 | 1 | 19,65 |
| | | | | 2 | 6,74 | | | | 2 | 29,99 |
| | | 2 | 7,02 | 3 | 2,89 | | 2 | 17,47 | 3 | 26,18 |
| | | | | 4 | 15,31 | | | | 4 | 6,85 |
| 5 | 307 | 1 | 78,84 | 1 | 109,30 | 541 | 1 | 25,25 | 1 | 27,85 |
| | | | | 2 | 38,25 | | | | 2 | 24,98 |
| | | 2 | 3,87 | 3 | 46,95 | | 2 | 0,15 | 3 | 41,87 |
| | | | | 4 | 32,17 | | | | 4 | 8,76 |
| 6 | 224 | 1 | 44,10 | 1 | 79,60 | 414 | 1 | 1,95 | 1 | 7,47 |
| | | | | 2 | 63,23 | | | | 2 | 42,20 |
| | | 2 | 44,73 | 3 | 61,85 | | 2 | 26,24 | 3 | 41,19 |
| | | | | 4 | 14,33 | | | | 4 | 9,82 |
| 7 | 293 | 1 | 6,04 | 1 | 54,03 | 373 | 1 | 8,15 | 1 | 32,22 |
| | | | | 2 | 40,65 | | | | 2 | 0,58 |
| | | 2 | 16,00 | 3 | 6,20 | | 2 | 9,15 | 3 | 14,09 |
| | | | | 4 | 16,42 | | | | 4 | 15,47 |

Таблица Г.3. Параметры распределений СП $N_i^{\{0,1\}}$, переданных «слонами» в каждый из дней

| i | $\bar{\mu}_i$ | k | $\frac{ \mu_{i,k} - \bar{\mu}_i }{\bar{\mu}_i}, \%$ | k | $\frac{ \mu_{i,k} - \bar{\mu}_i }{\bar{\mu}_i}, \%$ | $\bar{\sigma}_i$ | k | $\frac{ \sigma_{i,k} - \bar{\sigma}_i }{\bar{\sigma}_i}, \%$ | k | $\frac{ \sigma_{i,k} - \bar{\sigma}_i }{\bar{\sigma}_i}, \%$ |
|-----|---------------|-----|---|-----|---|------------------|-----|--|-----|--|
| 1 | 4039 | 1 | 1,34 | 1 | 0,35 | 798 | 1 | 7,32 | 1 | 10,84 |
| | | | | 2 | 2,24 | | | | 2 | 1,18 |
| | | 2 | 1,43 | 3 | 2,82 | | 2 | 5,69 | 3 | 13,54 |
| | | | | 4 | 5,82 | | | | 4 | 6,95 |
| 2 | 3143 | 1 | 1,41 | 1 | 1,14 | 724 | 1 | 5,73 | 1 | 10,01 |
| | | | | 2 | 4,05 | | | | 2 | 1,29 |
| | | 2 | 1,57 | 3 | 6,75 | | 2 | 7,39 | 3 | 12,52 |
| | | | | 4 | 3,64 | | | | 4 | 7,55 |
| 3 | 2818 | 1 | 2,77 | 1 | 8,05 | 595 | 1 | 1,37 | 1 | 5,19 |
| | | | | 2 | 2,01 | | | | 2 | 5,42 |
| | | 2 | 2,87 | 3 | 0,91 | | 2 | 2,30 | 3 | 1,25 |
| | | | | 4 | 6,60 | | | | 4 | 5,02 |
| 4 | 3739 | 1 | 12,88 | 1 | 19,79 | 1072 | 1 | 12,10 | 1 | 27,19 |
| | | | | 2 | 5,64 | | | | 2 | 4,21 |
| | | 2 | 12,78 | 3 | 16,75 | | 2 | 10,04 | 3 | 12,88 |
| | | | | 4 | 8,59 | | | | 4 | 8,24 |
| 5 | 3502 | 1 | 3,58 | 1 | 3,62 | 826 | 1 | 7,03 | 1 | 23,61 |
| | | | | 2 | 10,67 | | | | 2 | 1,72 |
| | | 2 | 3,63 | 3 | 3,98 | | 2 | 5,41 | 3 | 0,20 |
| | | | | 4 | 11,73 | | | | 4 | 5,96 |
| 6 | 1860 | 1 | 11,68 | 1 | 15,74 | 575 | 1 | 3,87 | 1 | 3,34 |
| | | | | 2 | 8,71 | | | | 2 | 8,55 |
| | | 2 | 11,16 | 3 | 14,93 | | 2 | 12,93 | 3 | 13,84 |
| | | | | 4 | 7,37 | | | | 4 | 13,25 |
| 7 | 1931 | 1 | 1,49 | 1 | 0,40 | 463 | 1 | 0,12 | 1 | 10,73 |
| | | | | 2 | 2,53 | | | | 2 | 11,32 |
| | | 2 | 1,52 | 3 | 5,59 | | 2 | 0,58 | 3 | 1,34 |
| | | | | 4 | 9,06 | | | | 4 | 6,67 |

Таблица Г.4. Параметры распределений СП $N_i^{\{1.0\}}$, переданных «мышами» в каждый из дней

| i | $\bar{\mu}_i$ | k | $\frac{ \mu_{i,k} - \bar{\mu}_i }{\bar{\mu}_i}, \%$ | k | $\frac{ \mu_{i,k} - \bar{\mu}_i }{\bar{\mu}_i}, \%$ | $\bar{\sigma}_i$ | k | $\frac{ \sigma_{i,k} - \bar{\sigma}_i }{\bar{\sigma}_i}, \%$ | k | $\frac{ \sigma_{i,k} - \bar{\sigma}_i }{\bar{\sigma}_i}, \%$ |
|-----|---------------|-----|---|-----|---|------------------|-----|--|-----|--|
| 1 | 96598 | 1 | 1,43 | 1 | 3,32 | 7370 | 1 | 3,71 | 1 | 9,42 |
| | | | | 2 | 0,44 | | | | 2 | 5,07 |
| | | 2 | 1,41 | 3 | 1,69 | | 2 | 0,83 | 3 | 12,03 |
| | | | | 4 | 4,48 | | | | 4 | 4,21 |
| 2 | 103816 | 1 | 0,08 | 1 | 0,15 | 6614 | 1 | 2,30 | 1 | 5,31 |
| | | | | 2 | 0,10 | | | | 2 | 1,46 |
| | | 2 | 0,17 | 3 | 0,98 | | 2 | 6,10 | 3 | 6,03 |
| | | | | 4 | 1,10 | | | | 4 | 14,06 |
| 3 | 100211 | 1 | 0,46 | 1 | 0,63 | 7887 | 1 | 3,85 | 1 | 5,07 |
| | | | | 2 | 1,65 | | | | 2 | 8,55 |
| | | 2 | 0,46 | 3 | 0,92 | | 2 | 0,55 | 3 | 10,10 |
| | | | | 4 | 0,05 | | | | 4 | 7,21 |
| 4 | 99411 | 1 | 4,19 | 1 | 7,52 | 9625 | 1 | 12,65 | 1 | 28,15 |
| | | | | 2 | 0,74 | | | | 2 | 10,23 |
| | | 2 | 4,40 | 3 | 3,67 | | 2 | 12,40 | 3 | 27,88 |
| | | | | 4 | 5,08 | | | | 4 | 2,92 |
| 5 | 97660 | 1 | 1,08 | 1 | 0,94 | 8924 | 1 | 5,48 | 1 | 8,88 |
| | | | | 2 | 1,18 | | | | 2 | 1,81 |
| | | 2 | 1,06 | 3 | 0,39 | | 2 | 5,49 | 3 | 6,71 |
| | | | | 4 | 1,74 | | | | 4 | 0,61 |
| 6 | 98499 | 1 | 0,17 | 1 | 0,36 | 10551 | 1 | 0,98 | 1 | 6,51 |
| | | | | 2 | 0,79 | | | | 2 | 3,07 |
| | | 2 | 0,20 | 3 | 0,41 | | 2 | 1,37 | 3 | 2,77 |
| | | | | 4 | 0,10 | | | | 4 | 2,92 |
| 7 | 89111 | 1 | 3,42 | 1 | 6,19 | 11970 | 1 | 9,22 | 1 | 10,72 |
| | | | | 2 | 0,65 | | | | 2 | 13,02 |
| | | 2 | 3,41 | 3 | 3,13 | | 2 | 2,36 | 3 | 0,73 |
| | | | | 4 | 10,04 | | | | 4 | 18,41 |

Таблица Г.5. Параметры распределений СП $N_i^{\{1.0\}}$, переданных «мулами» в каждый из дней

| i | $\bar{\mu}_i$ | k | $\frac{ \mu_{i,k} - \bar{\mu}_i }{\bar{\mu}_i}, \%$ | k | $\frac{ \mu_{i,k} - \bar{\mu}_i }{\bar{\mu}_i}, \%$ | $\bar{\sigma}_i$ | k | $\frac{ \sigma_{i,k} - \bar{\sigma}_i }{\bar{\sigma}_i}, \%$ | k | $\frac{ \sigma_{i,k} - \bar{\sigma}_i }{\bar{\sigma}_i}, \%$ |
|-----|---------------|-----|---|-----|---|------------------|-----|--|-----|--|
| 1 | 10550 | 1 | 5,82 | 1 | 2,87 | 3543 | 1 | 14,53 | 1 | 20,99 |
| | | | | 2 | 11,42 | | | | 2 | 15,71 |
| | | 2 | 1,98 | 3 | 14,32 | | 2 | 2,21 | 3 | 10,87 |
| | | | | 4 | 9,92 | | | | 4 | 25,70 |
| | | | | | | | | | | |
| | | | | | | | | | | |
| 2 | 11231 | 1 | 7,53 | 1 | 11,45 | 2949 | 1 | 0,18 | 1 | 3,05 |
| | | | | 2 | 4,19 | | | | 2 | 0,29 |
| | | 2 | 7,83 | 3 | 0,96 | | 2 | 8,54 | 3 | 9,37 |
| | | | | 4 | 15,10 | | | | 4 | 15,83 |
| | | | | | | | | | | |
| | | | | | | | | | | |
| 3 | 9648 | 1 | 2,59 | 1 | 0,09 | 2853 | 1 | 2,25 | 1 | 25,39 |
| | | | | 2 | 6,47 | | | | 2 | 18,33 |
| | | 2 | 2,11 | 3 | 0,58 | | 2 | 1,81 | 3 | 5,57 |
| | | | | 4 | 4,02 | | | | 4 | 3,71 |
| | | | | | | | | | | |
| | | | | | | | | | | |
| 4 | 7554 | 1 | 9,74 | 1 | 8,57 | 2891 | 1 | 24,63 | 1 | 16,14 |
| | | | | 2 | 13,99 | | | | 2 | 61,60 |
| | | 2 | 1,93 | 3 | 6,70 | | 2 | 11,73 | 3 | 30,55 |
| | | | | 4 | 11,74 | | | | 4 | 2,31 |
| | | | | | | | | | | |
| | | | | | | | | | | |
| 5 | 5358 | 1 | 12,06 | 1 | 25,30 | 3047 | 1 | 14,06 | 1 | 12,10 |
| | | | | 2 | 4,18 | | | | 2 | 20,10 |
| | | 2 | 31,77 | 3 | 11,21 | | 2 | 24,76 | 3 | 37,36 |
| | | | | 4 | 23,94 | | | | 4 | 45,99 |
| | | | | | | | | | | |
| | | | | | | | | | | |
| 6 | 3869 | 1 | 3,09 | 1 | 19,14 | 2416 | 1 | 22,56 | 1 | 35,54 |
| | | | | 2 | 1,70 | | | | 2 | 35,04 |
| | | 2 | 5,75 | 3 | 0,72 | | 2 | 13,66 | 3 | 38,53 |
| | | | | 4 | 6,14 | | | | 4 | 6,34 |
| | | | | | | | | | | |
| | | | | | | | | | | |
| 7 | 3649 | 1 | 0,73 | 1 | 23,25 | 2709 | 1 | 5,37 | 1 | 31,56 |
| | | | | 2 | 27,85 | | | | 2 | 2,95 |
| | | 2 | 7,85 | 3 | 12,37 | | 2 | 6,31 | 3 | 11,54 |
| | | | | 4 | 7,38 | | | | 4 | 18,25 |
| | | | | | | | | | | |
| | | | | | | | | | | |

Таблица Г.6. Параметры распределений СП $N_i^{\{1.0\}}$, переданных «слонами» в каждый из дней

| i | $\bar{\mu}_i$ | k | $\frac{ \mu_{i,k} - \bar{\mu}_i }{\bar{\mu}_i}, \%$ | k | $\frac{ \mu_{i,k} - \bar{\mu}_i }{\bar{\mu}_i}, \%$ | $\bar{\sigma}_i$ | k | $\frac{ \sigma_{i,k} - \bar{\sigma}_i }{\bar{\sigma}_i}, \%$ | k | $\frac{ \sigma_{i,k} - \bar{\sigma}_i }{\bar{\sigma}_i}, \%$ |
|-----|---------------|-----|---|-----|---|------------------|-----|--|-----|--|
| 1 | 40391 | 1 | 1,40 | 1 | 0,32 | 5581 | 1 | 0,91 | 1 | 1,98 |
| | | | | 2 | 2,28 | | | | 2 | 6,19 |
| | | 2 | 1,57 | 3 | 2,15 | | 2 | 1,27 | 3 | 5,94 |
| | | | | 4 | 5,85 | | | | 4 | 5,40 |
| | | | | | | | | | | |
| | | | | | | | | | | |
| 2 | 31427 | 1 | 1,34 | 1 | 2,41 | 5614 | 1 | 10,69 | 1 | 28,91 |
| | | | | 2 | 3,71 | | | | 2 | 3,51 |
| | | 2 | 1,77 | 3 | 6,74 | | 2 | 6,64 | 3 | 19,10 |
| | | | | 4 | 3,28 | | | | 4 | 4,46 |
| | | | | | | | | | | |
| | | | | | | | | | | |
| 3 | 28168 | 1 | 2,82 | 1 | 7,89 | 4618 | 1 | 0,82 | 1 | 2,50 |
| | | | | 2 | 1,86 | | | | 2 | 4,18 |
| | | 2 | 2,91 | 3 | 0,66 | | 2 | 3,64 | 3 | 1,06 |
| | | | | 4 | 6,57 | | | | 4 | 7,61 |
| | | | | | | | | | | |
| | | | | | | | | | | |
| 4 | 37355 | 1 | 12,85 | 1 | 19,95 | 8743 | 1 | 17,54 | 1 | 40,66 |
| | | | | 2 | 5,68 | | | | 2 | 13,69 |
| | | 2 | 12,73 | 3 | 17,15 | | 2 | 13,25 | 3 | 16,18 |
| | | | | 4 | 8,05 | | | | 4 | 10,87 |
| | | | | | | | | | | |
| | | | | | | | | | | |
| 5 | 34994 | 1 | 3,59 | 1 | 3,32 | 6419 | 1 | 10,50 | 1 | 28,74 |
| | | | | 2 | 10,82 | | | | 2 | 11,34 |
| | | 2 | 3,64 | 3 | 4,00 | | 2 | 4,83 | 3 | 8,82 |
| | | | | 4 | 11,20 | | | | 4 | 1,42 |
| | | | | | | | | | | |
| | | | | | | | | | | |
| 6 | 18572 | 1 | 12,56 | 1 | 15,54 | 4922 | 1 | 11,58 | 1 | 2,12 |
| | | | | 2 | 9,28 | | | | 2 | 13,80 |
| | | 2 | 11,29 | 3 | 15,04 | | 2 | 21,27 | 3 | 19,14 |
| | | | | 4 | 7,45 | | | | 4 | 25,75 |
| | | | | | | | | | | |
| | | | | | | | | | | |
| 7 | 19296 | 1 | 1,44 | 1 | 1,69 | 3747 | 1 | 3,58 | 1 | 4,21 |
| | | | | 2 | 2,68 | | | | 2 | 14,28 |
| | | 2 | 1,80 | 3 | 3,95 | | 2 | 2,99 | 3 | 11,03 |
| | | | | 4 | 9,43 | | | | 4 | 1,79 |
| | | | | | | | | | | |
| | | | | | | | | | | |

**Приложение Д. Результаты оценки параметров
распределений по каждому из дней, полученные по
полной выборке агрегированных последовательностей в
разрезе размера потоков**

Таблица Д.1. Параметры распределений СП $V_i^{\{0.1\}}$, созданных
«мышами» в каждый из дней

| i | $\bar{\mu}_i$ | k | $\frac{ \mu_{i,k} - \bar{\mu}_i }{\bar{\mu}_i}, \%$ | k | $\frac{ \mu_{i,k} - \bar{\mu}_i }{\bar{\mu}_i}, \%$ | $\bar{\sigma}_i$ | k | $\frac{ \sigma_{i,k} - \bar{\sigma}_i }{\bar{\sigma}_i}, \%$ | k | $\frac{ \sigma_{i,k} - \bar{\sigma}_i }{\bar{\sigma}_i}, \%$ |
|-----|-------------------|-----|---|-----|---|-------------------|-----|--|-----|--|
| 1 | $2,50 \cdot 10^6$ | 1 | 5,93 | 1 | 3,96 | $5,73 \cdot 10^5$ | 1 | 0,13 | 1 | 14,77 |
| | | | | 2 | 15,90 | | | | 2 | 6,14 |
| | | 2 | 5,79 | 3 | 4,71 | | 2 | 6,61 | 3 | 0,37 |
| | | | | 4 | 6,72 | | | | 4 | 15,47 |
| 2 | $2,86 \cdot 10^6$ | 1 | 4,64 | 1 | 9,22 | $5,61 \cdot 10^5$ | 1 | 11,30 | 1 | 2,06 |
| | | | | 2 | 0,61 | | | | 2 | 13,69 |
| | | 2 | 5,06 | 3 | 6,23 | | 2 | 19,52 | 3 | 21,68 |
| | | | | 4 | 3,75 | | | | 4 | 18,02 |
| 3 | $2,56 \cdot 10^6$ | 1 | 3,07 | 1 | 2,56 | $4,95 \cdot 10^5$ | 1 | 2,48 | 1 | 8,62 |
| | | | | 2 | 9,18 | | | | 2 | 14,12 |
| | | 2 | 3,11 | 3 | 0,35 | | 2 | 4,95 | 3 | 3,37 |
| | | | | 4 | 5,83 | | | | 4 | 8,07 |
| 4 | $2,57 \cdot 10^6$ | 1 | 2,71 | 1 | 7,54 | $5,01 \cdot 10^5$ | 1 | 2,68 | 1 | 4,19 |
| | | | | 2 | 2,20 | | | | 2 | 6,74 |
| | | 2 | 2,58 | 3 | 1,76 | | 2 | 3,65 | 3 | 1,36 |
| | | | | 4 | 3,28 | | | | 4 | 4,32 |
| 5 | $2,17 \cdot 10^6$ | 1 | 1,54 | 1 | 1,55 | $4,45 \cdot 10^5$ | 1 | 7,26 | 1 | 6,08 |
| | | | | 2 | 1,61 | | | | 2 | 8,93 |
| | | 2 | 1,22 | 3 | 0,50 | | 2 | 6,29 | 3 | 5,78 |
| | | | | 4 | 2,84 | | | | 4 | 9,16 |
| 6 | $2,25 \cdot 10^6$ | 1 | 1,50 | 1 | 3,23 | $5,75 \cdot 10^5$ | 1 | 12,43 | 1 | 30,69 |
| | | | | 2 | 0,28 | | | | 2 | 6,93 |
| | | 2 | 1,21 | 3 | 1,36 | | 2 | 12,41 | 3 | 11,53 |
| | | | | 4 | 1,06 | | | | 4 | 16,14 |
| 7 | $2,20 \cdot 10^6$ | 1 | 8,57 | 1 | 7,54 | $7,37 \cdot 10^5$ | 1 | 17,85 | 1 | 16,79 |
| | | | | 2 | 9,75 | | | | 2 | 21,78 |
| | | 2 | 9,73 | 3 | 15,45 | | 2 | 9,34 | 3 | 14,46 |
| | | | | 4 | 55,32 | | | | 4 | 21,27 |

Таблица Д.2. Параметры распределений СП $V_i^{\{0.1\}}$, созданных «мулами»
в каждый из дней

| i | $\bar{\mu}_i$ | k | $\frac{ \mu_{i,k} - \bar{\mu}_i }{\bar{\mu}_i}, \%$ | k | $\frac{ \mu_{i,k} - \bar{\mu}_i }{\bar{\mu}_i}, \%$ | $\bar{\sigma}_i$ | k | $\frac{ \sigma_{i,k} - \bar{\sigma}_i }{\bar{\sigma}_i}, \%$ | k | $\frac{ \sigma_{i,k} - \bar{\sigma}_i }{\bar{\sigma}_i}, \%$ |
|-----|-------------------|-----|---|-----|---|-------------------|-----|--|-----|--|
| 1 | $1,07 \cdot 10^6$ | 1 | 2,92 | 1 | 2,76 | $5,06 \cdot 10^5$ | 1 | 8,98 | 1 | 3,63 |
| | | | | 2 | 4,35 | | | | 2 | 18,48 |
| | | 2 | 2,64 | 3 | 0,82 | | 2 | 23,47 | 3 | 58,15 |
| | | | | 4 | 4,39 | | | | 4 | 11,45 |
| 2 | $1,25 \cdot 10^6$ | 1 | 7,99 | 1 | 10,11 | $4,80 \cdot 10^5$ | 1 | 2,18 | 1 | 1,81 |
| | | | | 2 | 5,20 | | | | 2 | 3,59 |
| | | 2 | 7,86 | 3 | 2,56 | | 2 | 7,07 | 3 | 6,46 |
| | | | | 4 | 13,51 | | | | 4 | 7,31 |
| 3 | $8,58 \cdot 10^5$ | 1 | 1,44 | 1 | 0,45 | $4,33 \cdot 10^5$ | 1 | 21,96 | 1 | 26,75 |
| | | | | 2 | 2,43 | | | | 2 | 6,18 |
| | | 2 | 1,48 | 3 | 0,36 | | 2 | 11,27 | 3 | 0,86 |
| | | | | 4 | 8,64 | | | | 4 | 5,77 |
| 4 | $8,47 \cdot 10^5$ | 1 | 7,93 | 1 | 9,84 | $4,51 \cdot 10^5$ | 1 | 0,07 | 1 | 2,84 |
| | | | | 2 | 9,81 | | | | 2 | 8,49 |
| | | 2 | 3,68 | 3 | 9,20 | | 2 | 13,12 | 3 | 0,89 |
| | | | | 4 | 17,91 | | | | 4 | 10,62 |
| 5 | $2,45 \cdot 10^5$ | 1 | 49,12 | 1 | 10,67 | $6,03 \cdot 10^5$ | 1 | 11,40 | 1 | 6,19 |
| | | | | 2 | 14,47 | | | | 2 | 5,62 |
| | | 2 | 37,06 | 3 | 28,98 | | 2 | 3,50 | 3 | 17,21 |
| | | | | 4 | 1,91 | | | | 4 | 24,76 |
| 6 | $3,66 \cdot 10^5$ | 1 | 15,77 | 1 | 19,35 | $4,01 \cdot 10^5$ | 1 | 17,02 | 1 | 22,20 |
| | | | | 2 | 3,03 | | | | 2 | 31,43 |
| | | 2 | 1,38 | 3 | 25,43 | | 2 | 6,76 | 3 | 1,79 |
| | | | | 4 | 69,81 | | | | 4 | 28,35 |
| 7 | $3,02 \cdot 10^5$ | 1 | 9,17 | 1 | 39,95 | $4,00 \cdot 10^5$ | 1 | 12,29 | 1 | 13,14 |
| | | | | 2 | 53,86 | | | | 2 | 19,56 |
| | | 2 | 37,69 | 3 | 44,42 | | 2 | 5,39 | 3 | 7,27 |
| | | | | 4 | 17,11 | | | | 4 | 4,09 |

Таблица Д.3. Параметры распределений СП $V_i^{\{0.1\}}$, созданных «слонами» в каждый из дней

| i | $\bar{\mu}_i$ | k | $\frac{ \mu_{i,k} - \bar{\mu}_i }{\bar{\mu}_i}, \%$ | k | $\frac{ \mu_{i,k} - \bar{\mu}_i }{\bar{\mu}_i}, \%$ | $\bar{\sigma}_i$ | k | $\frac{ \sigma_{i,k} - \bar{\sigma}_i }{\bar{\sigma}_i}, \%$ | k | $\frac{ \sigma_{i,k} - \bar{\sigma}_i }{\bar{\sigma}_i}, \%$ |
|-----|-------------------|-----|---|-----|---|-------------------|-----|--|-----|--|
| 1 | $4,93 \cdot 10^6$ | 1 | 1,11 | 1 | 0,81 | $1,04 \cdot 10^6$ | 1 | 11,69 | 1 | 17,40 |
| | | | | 2 | 3,19 | | | | 2 | 8,79 |
| | | 2 | 0,95 | 3 | 5,45 | | 2 | 12,29 | 3 | 26,24 |
| | | | | 4 | 7,27 | | | | 4 | 14,90 |
| 2 | $3,73 \cdot 10^6$ | 1 | 3,91 | 1 | 0,10 | $8,93 \cdot 10^5$ | 1 | 5,76 | 1 | 6,05 |
| | | | | 2 | 7,81 | | | | 2 | 3,71 |
| | | 2 | 3,69 | 3 | 8,93 | | 2 | 7,26 | 3 | 12,38 |
| | | | | 4 | 1,39 | | | | 4 | 8,37 |
| 3 | $3,09 \cdot 10^6$ | 1 | 8,60 | 1 | 16,20 | $7,86 \cdot 10^5$ | 1 | 5,95 | 1 | 5,52 |
| | | | | 2 | 1,21 | | | | 2 | 15,16 |
| | | 2 | 8,83 | 3 | 3,36 | | 2 | 5,01 | 3 | 3,67 |
| | | | | 4 | 14,12 | | | | 4 | 10,71 |
| 4 | $4,27 \cdot 10^6$ | 1 | 16,87 | 1 | 21,66 | $1,47 \cdot 10^6$ | 1 | 18,35 | 1 | 26,70 |
| | | | | 2 | 12,53 | | | | 2 | 11,10 |
| | | 2 | 16,58 | 3 | 22,66 | | 2 | 7,36 | 3 | 9,07 |
| | | | | 4 | 10,26 | | | | 4 | 8,18 |
| 5 | $3,93 \cdot 10^6$ | 1 | 5,14 | 1 | 4,34 | $1,06 \cdot 10^6$ | 1 | 6,11 | 1 | 25,80 |
| | | | | 2 | 14,55 | | | | 2 | 4,08 |
| | | 2 | 4,94 | 3 | 3,47 | | 2 | 3,23 | 3 | 2,28 |
| | | | | 4 | 15,00 | | | | 4 | 0,06 |
| 6 | $2,00 \cdot 10^6$ | 1 | 19,16 | 1 | 28,34 | $7,40 \cdot 10^5$ | 1 | 4,09 | 1 | 12,41 |
| | | | | 2 | 11,29 | | | | 2 | 7,83 |
| | | 2 | 17,71 | 3 | 20,80 | | 2 | 19,87 | 3 | 23,88 |
| | | | | 4 | 13,89 | | | | 4 | 15,58 |
| 7 | $2,20 \cdot 10^6$ | 1 | 1,30 | 1 | 0,77 | $5,38 \cdot 10^5$ | 1 | 1,54 | 1 | 12,75 |
| | | | | 2 | 3,14 | | | | 2 | 13,21 |
| | | 2 | 1,37 | 3 | 5,23 | | 2 | 1,09 | 3 | 2,10 |
| | | | | 4 | 8,34 | | | | 4 | 6,03 |

Таблица Д.4. Параметры распределений СП $V_i^{\{1.0\}}$, созданных «мышами» в каждый из дней

| i | $\bar{\mu}_i$ | k | $\frac{ \mu_{i,k} - \bar{\mu}_i }{\bar{\mu}_i}, \%$ | k | $\frac{ \mu_{i,k} - \bar{\mu}_i }{\bar{\mu}_i}, \%$ | $\bar{\sigma}_i$ | k | $\frac{ \sigma_{i,k} - \bar{\sigma}_i }{\bar{\sigma}_i}, \%$ | k | $\frac{ \sigma_{i,k} - \bar{\sigma}_i }{\bar{\sigma}_i}, \%$ |
|-----|-------------------|-----|---|-----|---|-------------------|-----|--|-----|--|
| 1 | $2,51 \cdot 10^7$ | 1 | 5,69 | 1 | 4,28 | $4,24 \cdot 10^6$ | 1 | 3,27 | 1 | 25,04 |
| | | | | 2 | 15,61 | | | | 2 | 9,21 |
| | | 2 | 6,01 | 3 | 5,07 | | 2 | 17,43 | 3 | 0,14 |
| | | | | 4 | 7,02 | | | | 4 | 30,81 |
| 2 | $2,87 \cdot 10^7$ | 1 | 4,46 | 1 | 9,29 | $3,84 \cdot 10^6$ | 1 | 15,06 | 1 | 11,78 |
| | | | | 2 | 11,38 | | | | 2 | 83,48 |
| | | 2 | 5,21 | 3 | 6,49 | | 2 | 25,79 | 3 | 31,95 |
| | | | | 4 | 4,06 | | | | 4 | 20,43 |
| 3 | $2,55 \cdot 10^7$ | 1 | 3,11 | 1 | 2,73 | $3,22 \cdot 10^6$ | 1 | 4,47 | 1 | 4,21 |
| | | | | 2 | 8,94 | | | | 2 | 18,36 |
| | | 2 | 3,08 | 3 | 0,28 | | 2 | 6,56 | 3 | 2,82 |
| | | | | 4 | 5,59 | | | | 4 | 19,23 |
| 4 | $2,57 \cdot 10^7$ | 1 | 2,53 | 1 | 7,47 | $3,35 \cdot 10^6$ | 1 | 5,25 | 1 | 7,07 |
| | | | | 2 | 2,32 | | | | 2 | 11,34 |
| | | 2 | 2,81 | 3 | 2,05 | | 2 | 7,72 | 3 | 5,24 |
| | | | | 4 | 3,44 | | | | 4 | 11,57 |
| 5 | $2,17 \cdot 10^7$ | 1 | 1,48 | 1 | 2,11 | $2,71 \cdot 10^6$ | 1 | 14,52 | 1 | 2,30 |
| | | | | 2 | 1,27 | | | | 2 | 15,34 |
| | | 2 | 1,52 | 3 | 3,10 | | 2 | 10,38 | 3 | 39,80 |
| | | | | 4 | 2,65 | | | | 4 | 12,95 |
| 6 | $2,24 \cdot 10^7$ | 1 | 1,54 | 1 | 2,46 | $4,70 \cdot 10^6$ | 1 | 19,74 | 1 | 40,61 |
| | | | | 2 | 0,07 | | | | 2 | 5,06 |
| | | 2 | 1,45 | 3 | 1,67 | | 2 | 19,00 | 3 | 17,62 |
| | | | | 4 | 2,75 | | | | 4 | 14,29 |
| 7 | $2,21 \cdot 10^7$ | 1 | 7,85 | 1 | 6,73 | $6,53 \cdot 10^6$ | 1 | 23,88 | 1 | 21,04 |
| | | | | 2 | 9,13 | | | | 2 | 23,90 |
| | | 2 | 9,73 | 3 | 14,09 | | 2 | 12,40 | 3 | 20,84 |
| | | | | 4 | 41,78 | | | | 4 | 15,77 |

Таблица Д.5. Параметры распределений СП $V_i^{\{1.0\}}$, созданных «мулами»
в каждый из дней

| i | $\bar{\mu}_i$ | k | $\frac{ \mu_{i,k} - \bar{\mu}_i }{\bar{\mu}_i}, \%$ | k | $\frac{ \mu_{i,k} - \bar{\mu}_i }{\bar{\mu}_i}, \%$ | $\bar{\sigma}_i$ | k | $\frac{ \sigma_{i,k} - \bar{\sigma}_i }{\bar{\sigma}_i}, \%$ | k | $\frac{ \sigma_{i,k} - \bar{\sigma}_i }{\bar{\sigma}_i}, \%$ |
|-----|-------------------|-----|---|-----|---|-------------------|-----|--|-----|--|
| 1 | $1,09 \cdot 10^7$ | 1 | 2,45 | 1 | 0,07 | $3,32 \cdot 10^6$ | 1 | 7,41 | 1 | 3,98 |
| | | | | 2 | 5,50 | | | | 2 | 11,56 |
| | | 2 | 2,40 | 3 | 9,59 | | 2 | 8,56 | 3 | 21,88 |
| | | | | 4 | 4,57 | | | | 4 | 11,26 |
| 2 | $1,26 \cdot 10^7$ | 1 | 8,08 | 1 | 9,38 | $3,44 \cdot 10^6$ | 1 | 5,74 | 1 | 5,62 |
| | | | | 2 | 5,30 | | | | 2 | 5,88 |
| | | 2 | 7,44 | 3 | 2,39 | | 2 | 13,32 | 3 | 14,82 |
| | | | | 4 | 12,99 | | | | 4 | 13,67 |
| 3 | $8,72 \cdot 10^6$ | 1 | 2,16 | 1 | 7,60 | $3,13 \cdot 10^6$ | 1 | 13,49 | 1 | 61,28 |
| | | | | 2 | 0,23 | | | | 2 | 1,57 |
| | | 2 | 2,77 | 3 | 1,09 | | 2 | 15,28 | 3 | 8,81 |
| | | | | 4 | 8,15 | | | | 4 | 18,29 |
| 4 | $8,58 \cdot 10^6$ | 1 | 12,29 | 1 | 9,68 | $3,07 \cdot 10^6$ | 1 | 17,55 | 1 | 14,38 |
| | | | | 2 | 37,00 | | | | 2 | 60,61 |
| | | 2 | 6,67 | 3 | 5,02 | | 2 | 0,45 | 3 | 22,49 |
| | | | | 4 | 18,24 | | | | 4 | 8,30 |
| 5 | $4,86 \cdot 10^6$ | 1 | 20,47 | 1 | 19,65 | $3,95 \cdot 10^6$ | 1 | 28,95 | 1 | 27,08 |
| | | | | 2 | 23,66 | | | | 2 | 35,49 |
| | | 2 | 35,87 | 3 | 9,31 | | 2 | 24,82 | 3 | 39,90 |
| | | | | 4 | 45,22 | | | | 4 | 49,39 |
| 6 | $3,17 \cdot 10^6$ | 1 | 22,69 | 1 | 38,42 | $3,67 \cdot 10^6$ | 1 | 9,57 | 1 | 4,91 |
| | | | | 2 | 29,09 | | | | 2 | 54,27 |
| | | 2 | 46,50 | 3 | 45,17 | | 2 | 40,38 | 3 | 49,65 |
| | | | | 4 | 0,41 | | | | 4 | 9,43 |
| 7 | $3,33 \cdot 10^6$ | 1 | 6,56 | 1 | 40,23 | $3,18 \cdot 10^6$ | 1 | 1,47 | 1 | 30,95 |
| | | | | 2 | 29,61 | | | | 2 | 10,80 |
| | | 2 | 17,45 | 3 | 15,63 | | 2 | 9,55 | 3 | 4,87 |
| | | | | 4 | 9,86 | | | | 4 | 17,66 |

Таблица Д.6. Параметры распределений СП $V_i^{\{1.0\}}$, созданных «слонами» в каждый из дней

| i | $\bar{\mu}_i$ | k | $\frac{ \mu_{i,k} - \bar{\mu}_i }{\bar{\mu}_i}, \%$ | k | $\frac{ \mu_{i,k} - \bar{\mu}_i }{\bar{\mu}_i}, \%$ | $\bar{\sigma}_i$ | k | $\frac{ \sigma_{i,k} - \bar{\sigma}_i }{\bar{\sigma}_i}, \%$ | k | $\frac{ \sigma_{i,k} - \bar{\sigma}_i }{\bar{\sigma}_i}, \%$ |
|-----|-------------------|-----|---|-----|---|-------------------|-----|--|-----|--|
| 1 | $4,94 \cdot 10^7$ | 1 | 0,88 | 1 | 1,08 | $7,33 \cdot 10^6$ | 1 | 8,81 | 1 | 14,47 |
| | | | | 2 | 2,90 | | | | 2 | 4,87 |
| | | 2 | 1,20 | 3 | 4,39 | | 2 | 9,98 | 3 | 20,27 |
| | | | | 4 | 7,40 | | | | 4 | 12,67 |
| 2 | $3,72 \cdot 10^7$ | 1 | 3,68 | 1 | 0,98 | $6,99 \cdot 10^6$ | 1 | 7,63 | 1 | 11,76 |
| | | | | 2 | 7,52 | | | | 2 | 5,47 |
| | | 2 | 4,04 | 3 | 8,88 | | 2 | 9,86 | 3 | 22,77 |
| | | | | 4 | 1,04 | | | | 4 | 7,96 |
| 3 | $3,10 \cdot 10^7$ | 1 | 8,39 | 1 | 16,16 | $6,31 \cdot 10^6$ | 1 | 10,26 | 1 | 12,64 |
| | | | | 2 | 1,46 | | | | 2 | 22,16 |
| | | 2 | 8,53 | 3 | 2,60 | | 2 | 8,10 | 3 | 0,08 |
| | | | | 4 | 13,75 | | | | 4 | 17,00 |
| 4 | $4,24 \cdot 10^7$ | 1 | 16,33 | 1 | 20,97 | $1,25 \cdot 10^7$ | 1 | 31,04 | 1 | 42,77 |
| | | | | 2 | 11,37 | | | | 2 | 23,11 |
| | | 2 | 17,50 | 3 | 23,37 | | 2 | 10,87 | 3 | 12,53 |
| | | | | 4 | 10,72 | | | | 4 | 13,01 |
| 5 | $3,93 \cdot 10^7$ | 1 | 5,10 | 1 | 4,49 | $8,14 \cdot 10^6$ | 1 | 7,86 | 1 | 26,13 |
| | | | | 2 | 14,05 | | | | 2 | 12,98 |
| | | 2 | 5,20 | 3 | 3,21 | | 2 | 0,65 | 3 | 5,67 |
| | | | | 4 | 14,03 | | | | 4 | 9,57 |
| 6 | $2,00 \cdot 10^7$ | 1 | 20,39 | 1 | 28,14 | $6,43 \cdot 10^6$ | 1 | 8,48 | 1 | 13,63 |
| | | | | 2 | 10,77 | | | | 2 | 15,15 |
| | | 2 | 17,32 | 3 | 20,83 | | 2 | 29,87 | 3 | 31,40 |
| | | | | 4 | 13,20 | | | | 4 | 28,80 |
| 7 | $2,19 \cdot 10^7$ | 1 | 1,21 | 1 | 0,98 | $4,21 \cdot 10^6$ | 1 | 1,05 | 1 | 15,50 |
| | | | | 2 | 3,27 | | | | 2 | 18,68 |
| | | 2 | 1,34 | 3 | 4,91 | | 2 | 3,74 | 3 | 2,11 |
| | | | | 4 | 7,91 | | | | 4 | 4,15 |

Приложение Е. Результаты вычисления параметров распределений случайных последовательностей, содержащих значения размеров пакетов

Таблица Е.1. Параметры ПР, созданных «мышами» в течение каждого из 7 дней

| i | $\bar{M}_{25,i}$ | k | $\frac{ M_{25,i,k} - \bar{M}_{25,i} }{\bar{M}_{25,i}}$ % | k | $\frac{ M_{25,i,k} - \bar{M}_{25,i} }{\bar{M}_{25,i}}$ % | $\bar{M}_{50,i}$ | k | $\frac{ M_{50,i,k} - \bar{M}_{50,i} }{\bar{M}_{50,i}}$ % | k | $\frac{ M_{50,i,k} - \bar{M}_{50,i} }{\bar{M}_{50,i}}$ % | $\bar{M}_{75,i}$ | k | $\frac{ M_{75,i,k} - \bar{M}_{75,i} }{\bar{M}_{75,i}}$ % | k | $\frac{ M_{75,i,k} - \bar{M}_{75,i} }{\bar{M}_{75,i}}$ % | $\bar{M}_{95,i}$ | k | $\frac{ M_{95,i,k} - \bar{M}_{95,i} }{\bar{M}_{95,i}}$ % | k | $\frac{ M_{95,i,k} - \bar{M}_{95,i} }{\bar{M}_{95,i}}$ % |
|-----|------------------|-----|---|-----|---|------------------|-----|---|-----|---|------------------|-----|---|-----|---|------------------|-----|---|-----|---|
| 1 | 66,54 | 1 | 50,26 | 1 | 50,26 | 76,72 | 1 | 94,76 | 1 | 96,65 | 102,89 | 1 | 183,71 | 1 | 130,01 | 1468,20 | 1 | 2,08 | 1 | 2,48 |
| | | | | 2 | 89,59 | | | | 2 | 180,04 | | | | 2 | 683,95 | | | | 2 | 4,26 |
| | | 2 | 43,70 | 3 | 126,73 | | 2 | 79,60 | 3 | 236,90 | | 2 | 113,05 | 3 | 325,02 | | 2 | 1,68 | 3 | 8,62 |
| | | | | 4 | 63,37 | | | | 4 | 117,50 | | | | 4 | 162,51 | | | | 4 | 2,28 |
| 2 | 97,08 | 1 | 25,46 | 1 | 7,49 | 143,61 | 1 | 37,46 | 1 | 11,14 | 233,75 | 1 | 39,81 | 1 | 24,26 | 1459,48 | 1 | 2,19 | 1 | 0,40 |
| | | | | 2 | 4,49 | | | | 2 | 7,09 | | | | 2 | 5,60 | | | | 2 | 0,60 |
| | | 2 | 13,48 | 3 | 23,96 | | 2 | 18,22 | 3 | 34,42 | | 2 | 17,42 | 3 | 32,97 | | 2 | 1,89 | 3 | 4,18 |
| | | | | 4 | 34,45 | | | | 4 | 50,62 | | | | 4 | 57,23 | | | | 4 | 2,79 |
| 3 | 72,36 | 1 | 26,12 | 1 | 24,11 | 88,35 | 1 | 47,72 | 1 | 42,79 | 126,16 | 1 | 57,63 | 1 | 49,56 | 1482,74 | 1 | 1,96 | 1 | 1,67 |
| | | | | 2 | 46,22 | | | | 2 | 80,64 | | | | 2 | 103,73 | | | | 2 | 3,53 |
| | | 2 | 4,02 | 3 | 28,13 | | 2 | 9,87 | 3 | 51,02 | | 2 | 25,36 | 3 | 72,61 | | 2 | 0,20 | 3 | 1,67 |
| | | | | 4 | 16,08 | | | | 4 | 31,27 | | | | 4 | 68,00 | | | | 4 | 0,78 |
| 4 | 95,62 | 1 | 9,12 | 1 | 3,04 | 137,79 | 1 | 13,72 | 1 | 6,33 | 214,85 | 1 | 17,60 | 1 | 6,77 | 1452,21 | 1 | 0,40 | 1 | 0,80 |
| | | | | 2 | 1,52 | | | | 2 | 1,06 | | | | 2 | 4,06 | | | | 2 | 0,40 |
| | | 2 | 15,21 | 3 | 16,73 | | 2 | 24,27 | 3 | 26,38 | | 2 | 31,13 | 3 | 36,54 | | 2 | 1,00 | 3 | 0,90 |
| | | | | 4 | 33,45 | | | | 4 | 50,65 | | | | 4 | 59,55 | | | | 4 | 2,60 |
| 5 | 66,54 | 1 | 2,19 | 1 | 30,59 | 75,27 | 1 | 3,86 | 1 | 57,95 | 89,81 | 1 | 8,10 | 1 | 90,67 | 1475,47 | 1 | 0,20 | 1 | 1,08 |
| | | | | 2 | 32,78 | | | | 2 | 61,82 | | | | 2 | 97,14 | | | | 2 | 2,37 |
| | | 2 | 39,33 | 3 | 34,96 | | 2 | 77,27 | 3 | 67,61 | | 2 | 126,28 | 3 | 116,57 | | 2 | 2,37 | 3 | 1,58 |
| | | | | 4 | 56,81 | | | | 4 | 108,18 | | | | 4 | 165,14 | | | | 4 | 4,24 |
| 6 | 101,44 | 1 | 1,43 | 1 | 31,53 | 153,78 | 1 | 1,89 | 1 | 44,44 | 275,92 | 1 | 7,90 | 1 | 42,68 | 1343,16 | 1 | 1,95 | 1 | 0,32 |
| | | | | 2 | 11,47 | | | | 2 | 16,07 | | | | 2 | 14,76 | | | | 2 | 5,85 |
| | | 2 | 11,47 | 3 | 30,10 | | 2 | 17,02 | 3 | 45,38 | | 2 | 30,04 | 3 | 73,25 | | 2 | 5,63 | 3 | 8,12 |
| | | | | 4 | 14,33 | | | | 4 | 20,80 | | | | 4 | 28,98 | | | | 4 | 3,14 |
| 7 | 69,45 | 1 | 29,31 | 1 | 50,25 | 82,54 | 1 | 59,90 | 1 | 100,41 | 335,53 | 1 | 199,77 | 1 | 191,10 | 1151,23 | 1 | 10,10 | 1 | 13,39 |
| | | | | 2 | 52,34 | | | | 2 | 102,17 | | | | 2 | 194,57 | | | | 2 | 9,73 |
| | | 2 | 18,84 | 3 | 508,74 | | 2 | 33,47 | 3 | 852,63 | | 2 | 45,93 | 3 | 242,67 | | 2 | 22,61 | 3 | 25,13 |
| | | | | 4 | 23,03 | | | | 4 | 40,52 | | | | 4 | 50,27 | | | | 4 | 22,73 |

Таблица Е.2. Параметры ПР, созданных «мулами» в течение каждого из 7 дней

| i | $\bar{M}_{25,i}$ | k | $\frac{ M_{25,i,k} - \bar{M}_{25,i} }{\bar{M}_{25,i}}$, % | k | $\frac{ M_{25,i,k} - \bar{M}_{25,i} }{\bar{M}_{25,i}}$, % | $\bar{M}_{50,i}$ | k | $\frac{ M_{50,i,k} - \bar{M}_{50,i} }{\bar{M}_{50,i}}$, % | k | $\frac{ M_{50,i,k} - \bar{M}_{50,i} }{\bar{M}_{50,i}}$, % | $\bar{M}_{75,i}$ | k | $\frac{ M_{75,i,k} - \bar{M}_{75,i} }{\bar{M}_{75,i}}$, % | k | $\frac{ M_{75,i,k} - \bar{M}_{75,i} }{\bar{M}_{75,i}}$, % | $\bar{M}_{95,i}$ | k | $\frac{ M_{95,i,k} - \bar{M}_{95,i} }{\bar{M}_{95,i}}$, % | k | $\frac{ M_{95,i,k} - \bar{M}_{95,i} }{\bar{M}_{95,i}}$, % |
|-----|------------------|-----|---|-----|---|------------------|-----|---|-----|---|------------------|-----|---|-----|---|------------------|-----|---|-----|---|
| 1 | 182,86 | 1 | 17,49 | 1 | 0,80 | 1398,41 | 1 | 0,52 | 1 | 6,24 | 1463,84 | 1 | 0,40 | 1 | 1,89 | 1503,10 | 1 | 0,10 | 1 | 0,39 |
| | | | | 2 | 6,36 | | | | 2 | 0,62 | | | | 2 | 0,20 | | | | 2 | 0,10 |
| | | 2 | 30,21 | 3 | 5,57 | | 2 | 1,25 | 3 | 0,31 | | 2 | 0,60 | 3 | 0,10 | | 2 | 0,10 | 3 | 0,00 |
| | | | | 4 | 148,69 | | | | 4 | 3,22 | | | | 4 | 1,49 | | | | 4 | 0,29 |
| 2 | 328,26 | 1 | 19,49 | 1 | 86,37 | 1423,13 | 1 | 1,12 | 1 | 4,80 | 1475,47 | 1 | 0,49 | 1 | 2,07 | 1506,00 | 1 | 0,10 | 1 | 0,39 |
| | | | | 2 | 19,93 | | | | 2 | 2,76 | | | | 2 | 1,18 | | | | 2 | 0,29 |
| | | 2 | 3,54 | 3 | 12,85 | | 2 | 0,82 | 3 | 3,17 | | 2 | 0,39 | 3 | 1,38 | | 2 | 0,10 | 3 | 0,29 |
| | | | | 4 | 51,82 | | | | 4 | 1,12 | | | | 4 | 0,59 | | | | 4 | 0,10 |
| 3 | 145,06 | 1 | 1,00 | 1 | 21,05 | 1354,79 | 1 | 3,22 | 1 | 2,47 | 1455,11 | 1 | 0,80 | 1 | 1,20 | 1501,64 | 1 | 0,19 | 1 | 0,19 |
| | | | | 2 | 30,07 | | | | 2 | 7,62 | | | | 2 | 2,70 | | | | 2 | 0,48 |
| | | 2 | 9,02 | 3 | 191,45 | | 2 | 8,37 | 3 | 41,96 | | 2 | 1,80 | 3 | 20,98 | | 2 | 0,29 | 3 | 4,07 |
| | | | | 4 | 25,06 | | | | 4 | 44,97 | | | | 4 | 4,90 | | | | 4 | 0,87 |
| 4 | 238,12 | 1 | 12,82 | 1 | 23,81 | 1428,94 | 1 | 1,42 | 1 | 2,95 | 1476,92 | 1 | 0,59 | 1 | 1,28 | 1506,00 | 1 | 0,10 | 1 | 0,29 |
| | | | | 2 | 17,10 | | | | 2 | 3,46 | | | | 2 | 1,38 | | | | 2 | 0,29 |
| | | 2 | 48,85 | 3 | 20,15 | | 2 | 0,61 | 3 | 3,87 | | 2 | 0,30 | 3 | 1,67 | | 2 | 0,10 | 3 | 0,29 |
| | | | | 4 | 405,46 | | | | 4 | 1,22 | | | | 4 | 0,69 | | | | 4 | 0,10 |
| 5 | 238,12 | 1 | 6,11 | 1 | 24,43 | 1379,51 | 1 | 7,17 | 1 | 6,64 | 1458,02 | 1 | 2,69 | 1 | 1,89 | 1501,64 | 1 | 0,48 | 1 | 0,29 |
| | | | | 2 | 57,40 | | | | 2 | 9,28 | | | | 2 | 3,89 | | | | 2 | 0,68 |
| | | 2 | 79,99 | 3 | 78,16 | | 2 | 1,69 | 3 | 1,37 | | 2 | 0,90 | 3 | 0,80 | | 2 | 0,10 | 3 | 0,10 |
| | | | | 4 | 93,43 | | | | 4 | 4,53 | | | | 4 | 2,09 | | | | 4 | 0,39 |
| 6 | 693,22 | 1 | 17,83 | 1 | 45,93 | 1402,77 | 1 | 0,62 | 1 | 2,28 | 1463,84 | 1 | 0,20 | 1 | 0,89 | 1503,10 | 1 | 0,00 | 1 | 0,19 |
| | | | | 2 | 52,86 | | | | 2 | 2,38 | | | | 2 | 0,89 | | | | 2 | 0,19 |
| | | 2 | 19,51 | 3 | 15,52 | | 2 | 0,10 | 3 | 3,73 | | 2 | 0,10 | 3 | 1,59 | | 2 | 0,00 | 3 | 0,29 |
| | | | | 4 | 56,21 | | | | 4 | 1,24 | | | | 4 | 0,60 | | | | 4 | 0,10 |
| 7 | 339,90 | 1 | 121,49 | 1 | 6,42 | 1359,15 | 1 | 2,25 | 1 | 2,78 | 1446,39 | 1 | 0,90 | 1 | 1,11 | 1500,19 | 1 | 0,10 | 1 | 0,19 |
| | | | | 2 | 214,32 | | | | 2 | 1,71 | | | | 2 | 0,50 | | | | 2 | 0,10 |
| | | 2 | 14,54 | 3 | 100,53 | | 2 | 2,46 | 3 | 0,64 | | 2 | 1,01 | 3 | 0,20 | | 2 | 0,19 | 3 | 0,00 |
| | | | | 4 | 24,38 | | | | 4 | 42,15 | | | | 4 | 20,51 | | | | 4 | 3,97 |

Таблица Е.3. Параметры ПР, созданных «слонами» в течение каждого из 7 дней

| i | $\bar{M}_{25,i}$ | k | $\frac{ M_{25,i,k} - \bar{M}_{25,i} }{\bar{M}_{25,i}}$ % | k | $\frac{ M_{25,i,k} - \bar{M}_{25,i} }{\bar{M}_{25,i}}$ % | $\bar{M}_{50,i}$ | k | $\frac{ M_{50,i,k} - \bar{M}_{50,i} }{\bar{M}_{50,i}}$ % | k | $\frac{ M_{50,i,k} - \bar{M}_{50,i} }{\bar{M}_{50,i}}$ % | $\bar{M}_{75,i}$ | k | $\frac{ M_{75,i,k} - \bar{M}_{75,i} }{\bar{M}_{75,i}}$ % | k | $\frac{ M_{75,i,k} - \bar{M}_{75,i} }{\bar{M}_{75,i}}$ % | $\bar{M}_{95,i}$ | k | $\frac{ M_{95,i,k} - \bar{M}_{95,i} }{\bar{M}_{95,i}}$ % | k | $\frac{ M_{95,i,k} - \bar{M}_{95,i} }{\bar{M}_{95,i}}$ % |
|-----|------------------|-----|---|-----|---|------------------|-----|---|-----|---|------------------|-----|---|-----|---|------------------|-----|---|-----|---|
| 1 | 1289,36 | 1 | 1,13 | 1 | 6,88 | 1421,67 | 1 | 0,51 | 1 | 3,38 | 1471,11 | 1 | 0,30 | 1 | 1,48 | 1504,55 | 1 | 0,10 | 1 | 0,29 |
| | | | | 2 | 1,01 | | | | 2 | 0,00 | | | | 2 | 0,00 | | | | 2 | 0,00 |
| | | 2 | 1,92 | 3 | 2,82 | | 2 | 1,02 | 3 | 2,45 | | 2 | 0,40 | 3 | 1,09 | | 2 | 0,10 | 3 | 0,19 |
| | | | | 4 | 8,80 | | | | 4 | 2,86 | | | | 4 | 1,28 | | | | 4 | 0,19 |
| 2 | 1203,57 | 1 | 1,57 | 1 | 4,11 | 1420,22 | 1 | 2,35 | 1 | 5,84 | 1475,47 | 1 | 1,28 | 1 | 2,96 | 1506,00 | 1 | 0,29 | 1 | 0,58 |
| | | | | 2 | 1,33 | | | | 2 | 3,38 | | | | 2 | 1,68 | | | | 2 | 0,39 |
| | | 2 | 22,71 | 3 | 32,50 | | 2 | 2,15 | 3 | 6,86 | | 2 | 0,99 | 3 | 3,15 | | 2 | 0,19 | 3 | 0,68 |
| | | | | 4 | 9,42 | | | | 4 | 2,76 | | | | 4 | 1,38 | | | | 4 | 0,29 |
| 3 | 283,19 | 1 | 30,29 | 1 | 25,67 | 1372,24 | 1 | 0,53 | 1 | 1,17 | 1453,66 | 1 | 0,30 | 1 | 0,20 | 1501,64 | 1 | 0,10 | 1 | 0,10 |
| | | | | 2 | 13,35 | | | | 2 | 3,50 | | | | 2 | 1,10 | | | | 2 | 0,19 |
| | | 2 | 232,07 | 3 | 39,53 | | 2 | 0,64 | 3 | 1,80 | | 2 | 0,50 | 3 | 0,80 | | 2 | 0,10 | 3 | 0,19 |
| | | | | 4 | 313,71 | | | | 4 | 1,06 | | | | 4 | 0,80 | | | | 4 | 0,19 |
| 4 | 427,14 | 1 | 55,15 | 1 | 1,02 | 1394,05 | 1 | 0,00 | 1 | 43,60 | 1460,93 | 1 | 0,30 | 1 | 21,30 | 1503,10 | 1 | 0,10 | 1 | 4,16 |
| | | | | 2 | 54,47 | | | | 2 | 1,56 | | | | 2 | 0,40 | | | | 2 | 0,10 |
| | | 2 | 190,29 | 3 | 199,48 | | 2 | 0,83 | 3 | 0,00 | | 2 | 0,50 | 3 | 0,20 | | 2 | 0,10 | 3 | 0,10 |
| | | | | 4 | 164,08 | | | | 4 | 4,07 | | | | 4 | 1,89 | | | | 4 | 0,39 |
| 5 | 341,35 | 1 | 17,46 | 1 | 54,52 | 1431,85 | 1 | 1,32 | 1 | 4,26 | 1482,74 | 1 | 0,69 | 1 | 2,06 | 1507,46 | 1 | 0,19 | 1 | 0,39 |
| | | | | 2 | 12,78 | | | | 2 | 2,64 | | | | 2 | 1,27 | | | | 2 | 0,29 |
| | | 2 | 19,17 | 3 | 69,86 | | 2 | 3,35 | 3 | 5,89 | | 2 | 1,57 | 3 | 2,94 | | 2 | 0,29 | 3 | 0,58 |
| | | | | 4 | 15,33 | | | | 4 | 6,40 | | | | 4 | 2,94 | | | | 4 | 0,58 |
| 6 | 320,99 | 1 | 59,34 | 1 | 43,49 | 1331,52 | 1 | 0,22 | 1 | 3,28 | 1436,21 | 1 | 2,73 | 1 | 0,30 | 1497,28 | 1 | 0,58 | 1 | 0,00 |
| | | | | 2 | 3,62 | | | | 2 | 1,97 | | | | 2 | 1,01 | | | | 2 | 0,29 |
| | | 2 | 200,21 | 3 | 127,74 | | 2 | 0,66 | 3 | 3,06 | | 2 | 0,51 | 3 | 1,52 | | 2 | 0,10 | 3 | 0,29 |
| | | | | 4 | 226,03 | | | | 4 | 1,42 | | | | 4 | 1,01 | | | | 4 | 0,19 |
| 7 | 796,45 | 1 | 2,92 | 1 | 16,25 | 1338,79 | 1 | 0,22 | 1 | 0,54 | 1434,76 | 1 | 0,10 | 1 | 0,30 | 1497,28 | 1 | 0,00 | 1 | 0,10 |
| | | | | 2 | 9,68 | | | | 2 | 1,63 | | | | 2 | 0,71 | | | | 2 | 0,10 |
| | | 2 | 2,92 | 3 | 45,46 | | 2 | 1,95 | 3 | 0,54 | | 2 | 0,81 | 3 | 0,41 | | 2 | 0,10 | 3 | 0,10 |
| | | | | 4 | 23,55 | | | | 4 | 4,13 | | | | 4 | 1,82 | | | | 4 | 0,29 |

Приложение Ж. Акты о внедрении



УРАЛЬСКИЙ БАНК ПАО «СБЕРБАНК РОССИИ»

620026, Екатеринбург, ул. Куйбышева, д.67
Телефон: 8 (800) 555-555-0
Факс: +7 (343) 354-62-90
info@uralsbank.ru, www.sberbank.ru

21.01.2016

№

на №

от

АКТ

об использовании результатов
кандидатской диссертационной работы
Копосова Александра Сергеевича

Настоящим актом подтверждаю, что комплексная методика оценивания параметров распределений случайных последовательностей с ограниченной областью рассеяния, разработанная в кандидатской диссертации А.С. Копосова, используется в Уральском Банке ПАО «Сбербанк России». Данная методика используется для анализа количества совершаемых транзакций на серверах системы приема платежей, с учетом результатов которого выбираются рекомендации по настройке алгоритмов балансировки и маршрутизации транзакций.

Управляющий директор ИТ
Уральского Банка ПАО «Сбербанк России»



С.Г. Смирнов

ОБЩЕСТВО С ОГРАНИЧЕННОЙ ОТВЕТСТВЕННОСТЬЮ
«МЕДИЦИНСКИЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ»

ОГРН: 1126686020317, ИНН: 6686016838, КПП: 668601001
 Юридический адрес: 624097, Свердловская обл., Верхняя Пышма г., Успенский пр., д. 125, корп. «Г», оф. 34.
 Р/С 40702810807030000973 в ЕФ АО "СМП БАНК", г. Екатеринбург, БИК 046577403, К/С 30101810865770000403

«УТВЕРЖДАЮ;»



Директор ООО «АйТи-Мед»

/А.О. Бобкова/

14 декабря 2015 г.

АКТ

об использовании результатов
 кандидатской диссертационной работы
 Александра Сергеевича Копосова

г. Верхняя Пышма

Комиссия в составе:

председатель В.В. Бобков; члены комиссии: В.В. Зюзин, Т.В. Голубятников,

составили настоящий акт о том, что результаты диссертационной работы «Разработка комплексной методики оценивания параметров функций распределения случайных последовательностей с ограниченной областью рассеяния с помощью генетических алгоритмов», представленной на соискание ученой степени кандидата технических наук по специальности «05.13.01 – Системный анализ, управление и обработка информации (связь и информатизация)», использованы в деятельности ООО «Медицинские информационные технологии (АйТи-Мед)» при разработке алгоритмов для программного экспертного медицинского диагностического комплекса «АвтоДиагност» в виде методики расчета при анализе статистических свойств кардио-интервалов.

Использование указанных результатов позволяет повысить быстродействие и надёжность алгоритма автоматического построения контуров левого желудочка сердца пациента на видеозаписях, получаемых со сканера УЗИ.

Результаты внедрялись при выполнении НИР и ОКР по теме «Экспертный программный медицинский диагностический комплекс для автоматического анализа видеоинформации, получаемой с УЗИ-сканеров», № гос.регистрации 01201366309.

Подписи:

Председатель комиссии:

Члены комиссии:

В.В. Бобков

В.В. Зюзин

Т.В. Голубятников

EiKONics

Общество с ограниченной ответственностью «Эйконикс»,
620085, г. Екатеринбург, ул. Аптекарская, 45-68.
ИНН 6679037121 КПП 667901001 ОГРН 1136679011699 ОКПО 26429551

АКТ

об использовании результатов, полученных в
диссертационном исследовании
Копосова Александра Сергеевича

Настоящим актом подтверждаю, что комплексная методика оценивания параметров распределений случайных последовательностей с ограниченной областью рассеяния, разработанная в кандидатской диссертации А.С. Копосова, используется в ООО «Эйконикс» в информационно-измерительной системе для автоматизации процессов классификации, контроля качества и учета лесоматериалов.

Директор ООО «Эйконикс»



А.В. Круглов

19.01.2016



Филиал Акционерного общества «ЭР-Телеком Холдинг» в городе Екатеринбург
(Филиал в г.Екатеринбург АО «ЭР-Телеком Холдинг»)

20 января 2016

№ _____

г.Екатеринбург



Утверждаю:
Директор филиала
АО «ЭР-Телеком Холдинг»
Э.В. Афонцев

АКТ

об использовании результатов
кандидатской диссертационной работы
Копосова Александра Сергеевича

Комиссия в составе председателя и членов комиссии составили настоящий акт в том, что комплексная методика оценивания параметров распределений случайных последовательностей с ограниченной областью рассеяния используется в екатеринбургском филиале АО «ЭР-Телеком Холдинг». Данная методика используется для анализа статистических свойств Интернет-трафика, с учетом результатов которого выбираются рекомендации для настройки сетевого оборудования и параметры режимов обеспечения скорости конечных пользователей для обеспечения гарантированного качества обслуживания пользователей сети Интернет.

Комиссия в составе:

Председатель комиссии:

Технический директор

Михайлов А.В.

Члены комиссии:

Руководитель ЦГС

Благинин А.В.

Руководитель ОТСК

Масленников М.В.

Приложение 3. Свидетельство о регистрации программы для ЭВМ

РОССИЙСКАЯ ФЕДЕРАЦИЯ



СВИДЕТЕЛЬСТВО

о государственной регистрации программы для ЭВМ

№ 2016614275

«Программная библиотека ES&RP»

Правообладатель: *Федеральное государственное автономное образовательное учреждение высшего профессионального образования «Уральский федеральный университет имени первого Президента России Б.Н. Ельцина» (RU)*

Авторы: *Копосов Александр Сергеевич (RU),
Поршнев Сергей Владимирович (RU)*



Заявка № 2016611747

Дата поступления 02 марта 2016 г.

Дата государственной регистрации

в Реестре программ для ЭВМ 20 апреля 2016 г.

Руководитель Федеральной службы
по интеллектуальной собственности

Г.П. Ивлиев